THE UNIVERSITY
of EDINBURGH

Natural Products: Biosynthesis, Antimicrobial Properties and Protein Targets

Daynea Wallock-Richards

Thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

2015

Abstract

The diversity of biosynthetic pathways in prokaryotes and eukaryotes has led to numerous bioactive natural products (NPs) which occupy a vast chemical space. Despite the current challenges in NP research, these molecules are still relevant today as they are a major source of human medicine as well as being useful biological tools. The elucidation of their biosynthetic pathways has also provided information about the biochemical and biophysical properties of fascinating enzyme families such as the α-oxoamine synthases (AOSs). The AOSs are an expanding group of pyridoxal 5'-phosphate (PLP)-dependent enzymes, which are involved in the biosynthesis of several important NP, including those essential for life. This study reports the characterization and structural analysis of a unique AOS denoted as TamD from *Pseudoalteromonas tunicata*. This enzyme plays a major role in tambjamine biosynthesis and consists of both an acyl carrier protein (ACP) domain and a PLP-binding catalytic domain. UV/vis spectroscopy and mass spectrometry (MS) of the recombinant TamD purified from *E. coli* revealed that the enzyme forms a Schiff base with PLP via Lys380, which is responsible for its characteristic yellow colour. It binds L-serine as a natural substrate with a K_d of 5.01 ± 0.64 mM. This thesis also reports structural data for TamD from x-ray crystallography at a resolution of 4.98 Å, which shows four molecules in the asymmetric unit (ASU) suggesting the enzyme exist as a dimer. The absence of the N-terminal region where the ACP domain is located in the crystal strucuture also suggests intrinsic flexibility and disorder within that region.

With the increasing demand for new anti-infective therapies, investigations of the molecular interactions between NPs and their protein targets are vital in understanding the inhibition or activation properties of these molecules. The cysteine transpeptidases known as sortases produced by Gram positive bacteria have been identified as attractive targets for NP inhibitors. In this thesis, the molecular basis for the inhibition of *Streptococcus mutans* sortase A (SrtA) by the plant flavonoid, *trans*-chalcone is explored, using a combination of MS, enzyme kinetics, molecular modelling and x-ray

crystallography. This study reports the first high resolution crystal structure of the H139A mutant of *S. mutans* SrtA, which reveals a unique N-terminal α-helix domain. *Trans*-chalcone was found to inhibit the *in vitro* activity of *S. mutans* SrtA in a slow, tight–binding manner, with a half maximal inhibitory concentration (IC_{50}) of 5.0 ± 0.6 μM. The interaction resulted in a covalent adduct with the active site cysteine residue (Cys205) via a Michael addition mechanism. Additionally, *trans*-chalcone showed evidence of *S. mutans* anti-biofilm activity in a concentration dependent manner up to 250 μM with an efficacy cut-off point at higher concentations. These results indicate that chalcone flavonoids are worth further investigation as potential antibiofilm inhibitors.

A renewed interest in plant NPs has also led to a collaborative investigation on the antimicrobial potential of garlic-derived allicin, against *Burkholderia cepacia* complex (Bcc), the major bacterial phytopathogen for alliums and an intrinsically multiresistant and life-threatening human pathogen. Allicin is the principal antibacterial agent in fresh preparations of garlic extracts. This investigation reports the first evidence that allicin and allicin-contaning garlic extracts possess inhibitory and bactericidal activities against Bcc. The minimum inhibitory concentrations (MICs) of aqueous garlic extract (AGE) against 38 Bcc isolates ranged from 0.5 to 3% (v/v). An investigation into the possible molecular mechanisms of allicin with a recombinant thiol-dependent peroxiredoxin (BCP) from *B. cenocepacia* revealed that allicin and AGE modify an essential BCP catalytic cysteine residue and suggests a role for allicin as a general electrophilic reagent that targets protein thiols. Present therapeutic options against these life-threatening pathogens are limited; thus, allicin-containing compounds merit further investigation as adjuncts to existing antibiotics.

Acknowledgments

I would like to take this opportunity to acknowledge all the persons who have contributed to the completion of this thesis and who have supported me throughout my studies.

I would first like to thank my supervisor, Dr. Dominic Campopiano for giving me the opportunity to work in his research group. His continuous words of encouragement, insightful discussions, academic advice, financial support and contagious passion for science are very much appreciated.

In extension, I must also thank the University of Ediburgh scholarship selection team and the Tercentenary Committee at the School of Chemistry for the financial contributions I have received from the scholarship awards. My academic endeavours at Edinburgh would not have been possible without this financial support.

I would like to express my gratitude to Professor John Govan and Dr. Cathy Doherthy who have given me the privilege of collaborating with them on the Allicin project. They have been very generous in providing part of my funding during a difficult period of my studies. It has also been a rewarding experience learning from such brilliant minds.

I would like to thank Professor Bryan Hanley, who believed in the sortase project from the onset thus making the Edinburgh-Chicago collaboration possible. Through this partnership, I was able to work with Dr. Mike Dodds and Dr. Amarnath Maitra from Wrigley's. They have not only provided me with the financial support needed to carry out this project, but they have also been very generous in hosting me at the Global Innovative Centre during my summer internship, where I was able to learn new and useful scientific skills which contributed immensely to this project. Therefore, I would like to express my sincerest gratitude to them for their contributions.

I am indebted to Dr. Dave Clarke for his help with mass spectrometry analysis and also for his astute academic advice and suggestions during my studies. Many thanks also to Dr. Jon Marles-Wright for his collaboration and help with structural analysis. His dedication in sharing his scientific expertise in the sortase project is very much appreciated.

My sincerest gratitude goes to Professor Jim Naismith and Dr. Stephen McMahon for their collaborative efforts on the TamD project. They have given me the opportunity to work in their laboratory in St. Andrews where I felt at home. They have provided useful suggestions and scientific expertise in x-ray crystallography which have no doubt assisted greatly in my work.

My heart-felt appreciation goes to the past and present members of the Campopiano group (Lab 229) for their support and friendship during my time at Edinburgh. Special thanks to Dr. Jonathan Lowther, who patiently and willingly taught me the survival skills I needed during my induction to the lab. Special thanks to Ashley Emily Beattie for her continued friendship and encouragement through the tough days. Lily, Chris and Guiomar thanks a lot for the fruitful discussions we have had, that have aided in my scientific growth.

I would like to say a big thank you to my dearest family for their love, their prayers, their encouragement and belief in me. My husband, mom, dad and sisters, you have truly been my strength throughout my studies.

Declaration

I, Daynea Wallock-Richards, hereby certify that this thesis has been composed by myself. This document is a record of my work, and has not been accepted in partial or complete fulfilment of any other degree or professional qualification.

Daynea Wallock-Richards
The University of Edinburgh
2015

Table of Contents

Abbreviations

4’-PP	4’-phosphopantetheinyl
A_{100}	Ampicillin (100μg/mL)
AAA	ACTA active attachment
AAS	Aqueous allicin standard
AAT	Aspartate aminotransferase
Abs	Absorbance
ACP	Acyl carrier protein
AcpS	Acyl carrier protein synthase
AGE	Aqueous garlic extract
AKB	α-amino-β-ketobutyrate
ALA	5-aminolevulinic acid
ALAS	Aminolevulinate synthase
AON	8-amino-7-oxononanoate
AONS	8-Amino-7-oxononanoate synthase
AOS	α-Oxoamine synthase
APS	Ammonium persulfate
ASU	Asymmetric unit
ATP	Adenosine triphosphate
AT	Acyltransferase
Bcc	*Burkholderia cepacia* complex
B. cenocepacia	*Burkholderia cenocepacia*
BCP	Bacterioferritin comigratory protein
BHI	Brain heart infusion broth
BLAST	Basic local alignment search tool
Bp	base pairs
C4H	Cinnamate 4-hydroxylase
CAI-1	Cholera autoinducer-1
CF	Cystic fibrosis
CFE	Cell free extract
CFU	Cell forming units
CHS	Chalcone synthase
CID	Collision induced dissociation
CLF	Chain lengthening factor
CoA	Coenzyme A
CsqA	Cholera CAI-1 synthase
CV	Column volume
Dabcyl	4-([4-(dimethylamino) phenyl]azo)-benzoyl
DH	Dehydratase

DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E. coli	*Escherichia coli*
ECD	Electron capture dissociation
Edans	[(2-aminoethyl)-amino]naphthalene-1-sulphonyl
EDTA	Ethylenediaminetetraacetic acid
Ex/Em	Excitation/Emission
ER	Enoylreductase
FAD	Flavin adenine dinucleotide
FAS	Fatty acid synthase
FDA	Food and drug administration
FRET	Förster resonance energy transfer
FT-ICR-ESI	Fourier transform ion cyclotron resonance electron spray ionization
HBM	4-hydroxy-2-2'-bipyrrole-5-methanol
HSN1	Hereditary sensory neuropathy type 1
IPTG	Isopropylthio-β-galactoside
k_{cat}	Catalytic constant
K_d	Dissociation constant
KBL	2-amino-3-ketobutyrate CoA ligase
Kbp	Kilobase pair
Km	Michaelis constant
KPhos	Potassium phosphate
KR	Ketoreductase
KS	Ketosynthase
LAI-1	*Legionella* autoinducer-1
LB	Luria Bertani
LC-MS	Liquid chromatography-mass spectrometry
LMW	Low molecular weight marker
L. pneumophila	*Legionella pneumophila*
L-Ser	L-serine
Lys	Lysine
MBC	4-methoxy-2-2'-bipyrrole-5-carbaldehyde
MHA	Muller-Hinton agar
MHB	Muller-Hinton broth
MRSA	Methicillin-resistant *Staphylococcus aureus*
MS	Mass spectrometry
NCBI	National Centre for Biotechnology Information

NCCLS	National Committee for Clinical Laboratory Standards
Ni-NTA	Nickel-nitrilotriacetic acid
NMR	Nuclear magnetic resonance
NP	Natural products
NRPS	Nonribosomal peptide synthase
OD_{600}	Optical density at 600 nm
ORF	Open reading frames
P. tunicata	*Pseudoalteromonas tunicata*
PAL	Phenylalanine ammonia lyase
PCP	Peptide carrier protein
PCR	Polymerase chain reaction
PDB	Protein data bank
pI	Isoelectric point
PKS	Polyketide synthase
PLP	Pyridoxal 5'-phosphate
PNP	Plant natural products
PPTase	Phophopantetheinyl transferase
RP-HPLC	Reverse phase high performance liquid chromatography
ROS	Reactive oxygen species
S. aureus	*Staphylococcus aureus*
S. coelicolor	*Streptomyces coelicolor*
S. mutans	*Streptococcus mutans*
S. pyogenes	*Streptococcus pyogenes*
SAM	(*S*)-adenosylmethionine
SEC	Size exclusion chromatography
SDM	Site directed mutagenesis
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
S-Gal	3,4-cyclohexenoesculetin- β-D-galactopyranoside
SOC	Super optimal broth with catabolite repression
SPT	Serine palmityoltransferase
SrtA	Sortase A
TAE	Tris base, acetic acid and EDTA
TEV	Tobacco etch virus
TGS	Tris-Glycine-SDS
Tris-HCl	2-Amino-2-hydroxymethyl-1,3-propanediol-Hydrochloric acid
UV/vis	Ultraviolet/visible
V. cholera	*Vibrio cholerae*
Vmax	Maximum velocity
WT	Wild type

Chapter 1: General Introduction to Natural Products

PART I: Biosynthesis of Antibacterial Natural Products

1.1 Why study Natural Products?

1.1.1 Pathways and Products

Natural products (NPs) are low molecular weight organic molecules made by bacteria, fungi, lichens, marine invertebrates and plants. Their revolutionary applications in the pharmaceutical, agricultural, food and cosmetic industries cannot be underestimated.[1-3] Traditionally, NPs have been a major source of new drugs, which led to approximately 80% of all drugs by 1990 being NPs and NP derivatives.[4] However, over the last two decades modern drug discovery efforts have been more focused on screening synthetic combinatorial libraries,[5] resulting in a decline in the use of NPs in the pharmaceutical industry. This trend may lead one to question if there is still a role for NPs in the discovery of novel compounds and new chemical processes. Whatever doubts the current reader might have, it is without reservation that I do propose that NPs are still relevant today and remain a source of inspiration for chemists and biologists. There is untapped potential in undiscovered compounds from unexplored biological sources that are now more accessible, through the advances in synthetic biology, genomics and proteomics which allow for manipulation of biosynthetic pathways.[6, 7] NPs are also a vital source for identifying lead molecules for drug development, as over half of all Food and Drug Administration (FDA)-approved drugs are either NPs, semisynthetic derivatives of NP scaffolds or synthetic compounds with a NP pharmacophore.[4] Current research has also shown that NPs serve as important chemical tools for identifying and understanding the mechanisms of pharmacologically relevant biological targets.[1] The use of NPs as enzyme inhibitors or activators has made it possible for us to understand complex interactions of many protein targets, as well as identify novel proteins as targets for new classes of NPs.

Biosynthetic pathways are nature's blueprints/route maps to diverse classes of NPs and an understanding of these pathways is crucial for the discovery of new biologically

active secondary metabolites. The focus of NP discovery is no longer solely on the end products, but also, on the biochemical processes involved in the transformation from genes to molecules.[1] Biosynthetic pathways consist of genes that encode for several proteins that play a role in the enzymatic assembly of NPs. They were classically elucidated through feeding studies and radioactive labelling experiments up until the 1980s.[1] Pioneer work on cloning the biosynthetic gene cluster for the antibiotic actinorhodin[8] and later erythromycin[9, 10] have paved the way for further molecular studies of polyketide synthases and their ancillary enzymes. The study of biosynthetic pathways responsible for the production of polyketides (PKs), non-ribosomal peptides (NRPs) and fatty acids (FAs) have resulted in the emergence of several classes of enzymes with diverse functions. What is common about these pathways is the use of a combination of enzymes from different classes to process covalent acyl-enzyme intermediates, iterative cycles of condensation and chemical tailoring of monomer units and intermediates to produce a diverse array of NPs.[11] With automated bioinformatics platforms, it is now easier to predict and identify gene clusters which encode for novel NPs and study the function of fascinating enzymes which carry out interesting chemical reactions along these pathways.[6, 12]

1.1.2 Antibacterial Resistance

Infectious diseases remain the second-leading cause of death worldwide, with bacterial infections claiming the lives of over 17 million globally. Of particular concern is the emergence of bacterial strains with resistance to current and newly developed antibiotics.[13, 14] Since the discovery and use of penicillin and other β-lactams, the rise in antimicrobial resistance has been linked to the overuse and misuse of antibiotics through selection pressure. Despite this problem, the demand for the development of novel antimicrobial therapies has outweighed the rate at which new therapies are being developed and approved for clinical use.[15] Therefore a renewed interest in NPs as leads for antibiotic discovery coupled with the advances in genomics and sequencing technologies may be one of the best approaches to address the current crisis. NPs are

regarded as superior over synthetic compounds for antibiotic discovery as these molecules have evolved to cope with microbial biology, which makes them more likely to penetrate the bacterial cell membrane and interact with specific protein targets, in comparison to their synthetic counterparts. NPs also occupy a more diverse chemical space and offer more structural complexity which is required for the inhibition of many antibacterial protein targets, making them suitable candidates for continuous antibiotic discovery.[4, 13]

1.1.2.1 *Burkholderia cepacia complex (Bcc)*

In the last few decades, *Burkholderia cepacia* complex (Bcc), a group of 17 closely-related species distributed widely in soil, water and the plant rhizosphere[16] has emerged as an important opportunistic human pathogen, in particular as a cause of life-threatening lung infections in individuals with cystic fibrosis (CF) and chronic granulomatous disease.[17, 18] Unfortunately, a common feature of the Bcc is its intrinsic resistance to most antibiotics[19]; hence antibiotic treatment presents a major challenge. Treatment of Bcc infections remains challenging and no particular antibiotic regimen is presently recognised.[20, 21] Although patient segregation and strict infection control have reduced the incidence of Bcc infections in individuals with CF, such infections remain an important clinical problem. At present, the most predominant Bcc species responsible for CF infections are *B. cenocepacia* and *B. multivorans*.[17, 22] Most transplant centres worldwide exclude individuals infected by *B. cenocepacia* from access to lung transplantation, the only proven treatment for severe CF lung disease. There is an even greater need to have alternative therapies with the recent alarming report of a clinical trial of inhaled aztreonam that failed to treat CF patients with *Burkholderia spp.* infection.[23, 24] Thus, the availability of novel antimicrobial agents against Bcc would represent a major clinical advance for patients with Bcc infection. Previous investigations of novel NPs against Bcc species have included mushroom extracts[25], polyketides[26], docosahexaenoic omega-3 fatty acid[27] and microbe-derived volatile organic compounds.[28] Given the dearth of new antibiotics from conventional approaches

within the pharmaceutical industry, the development of novel antimicrobial compounds derived from natural sources would be welcomed[29, 30]; this is particularly relevant to the management of lung infections which are the primary cause of morbidity and mortality in the CF population.

1.2 Tambjamines

1.2.1 Chemical and Biological Properties

Tambjamines are NPs isolated from bacteria and marine invertebrates including bryozoans, nudibranchs and ascidians. They belong to the family of 4-methoxypyrrolic NPs, which includes the prodiginines[31, 32] e.g. prodigiosin, undecylprodigiosin and streptorubin B[33] (Figure 1). They are structurally characterized by a 2,2'-bipyrrole ring system, with ring B having an enamine moiety at the C5 position and a methoxy group at the C3 position (Figure 1). The enamine nitrogen usually displays a short alkyl side chain, which contributes to diversity within the family. Additionally, some tambjamines have a bromide substituent at the C5 position of ring A.[34, 35] Tambjamines A-D were the first members of the family to be isolated from nudibranchs in 1982 by Carté and Faulkner, during their studies of defense mechanisms of marine invertebrates.[36] Tambjamines were later reported to be produced by the terrestrial bacterium *Streptomyces* BE186591.[37, 38] More recently, the marine bacterium *Pseudoalteromonas tunicata* was also identified as a producer of tambjamines.[39]

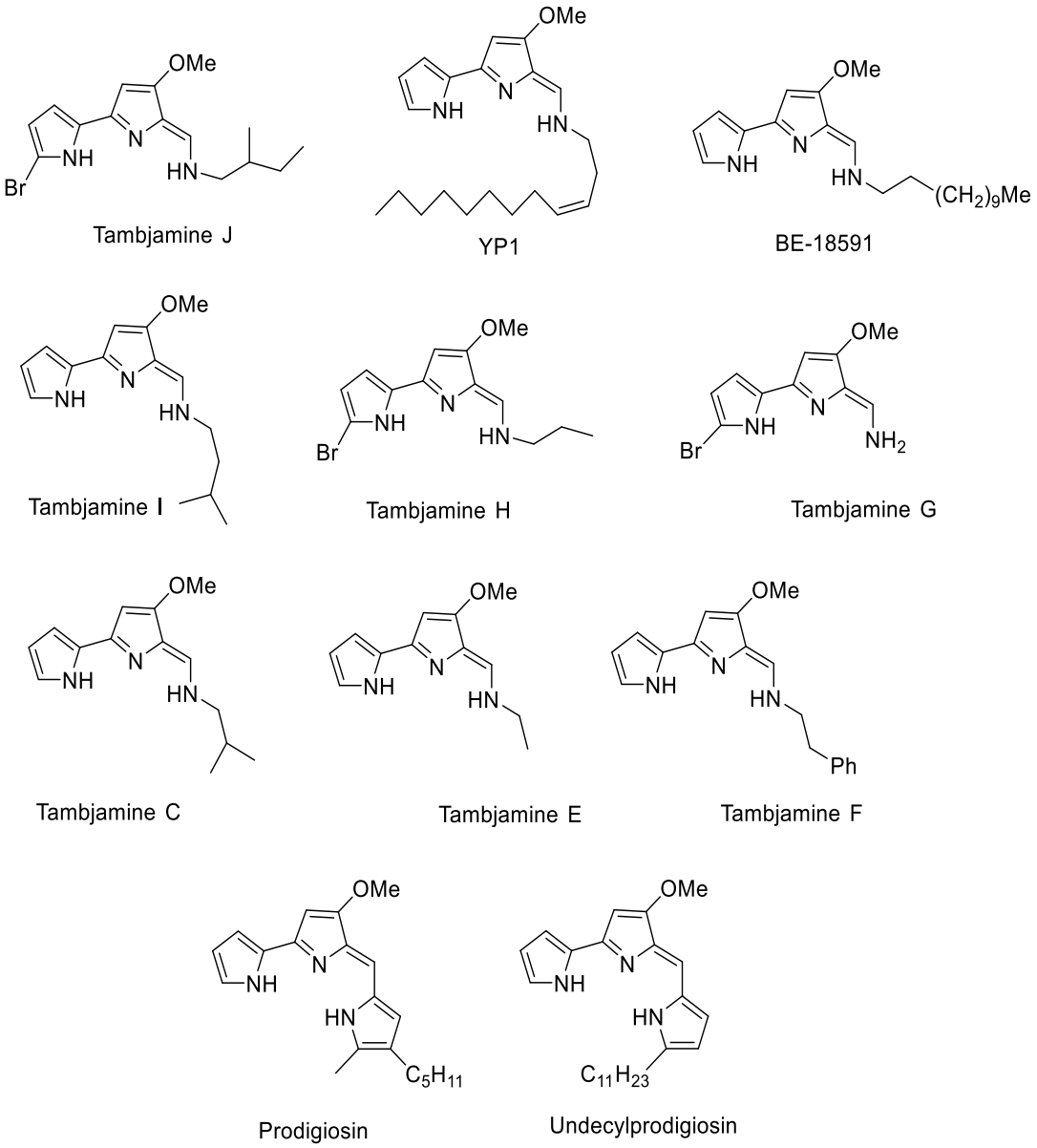


Figure 1: Chemical structures of tambjamines and the prodiginines, prodigiosin and undecylprodigiosin. Tambjamines are bipyrrolic natural product isolated from *Streptomyces sp.* and the marine bacterial *Pseudoalteromonas sp.* whereas the prodiginines are tripyrrolic natural products produced by *Serratia* and *Streptomyces sp*.

Tambjamine YP1 isolated from *P. tunicata* is the only member of the family identified so far with an unsaturated alkyl side chain. The lipophilic nature of the alkyl side chain has been credited for the increased bioactivity of YP1 in comparison to other tambjamines.[40] Tambjamines are pharmacologically important molecules as they possess antimicrobial, anticancer and immunosuppressive properties. These properties are partly due to their ability to intercalate DNA and participate in the oxidative cleavage of single-strand DNA.[35] The latest research on the biological activity of tambjamines has shown that they have good ionophoric properties. Tambjamines can move across cellular membranes, disturb the normal ionic balance and alter the intracellular pH, which can trigger apoptosis.[41, 42] Ecologically, tambjamines provide a natural defense to marine invertebrates against predators. Also, the antifouling property of tambjamines produced by *P. tunicata* has made this organism a widely studied model for marine biofilm formation studies.[43] Since these molecules possess pharmacological properties, their pathways are worth further investigation in the quest to find new antibiotic leads.

1.2.2 Biosynthesis

Subsequent to the isolation of tambjamine YP1, the gene cluster involved in its biosynthesis was identified by Burke *et. al.* in 2007. So far, this has been the only report of the identification of the biosynthetic genes (tam cluster) for the tambjamine class of compounds, since their first isolation in 1982.[39] This significant step in tambjamine research has led to the cloning and expression of the *tam* cluster in *E. coli* to produce a yellow tambjamine product, identified as YP1. The *tam* cluster is 34 kb(s) in length and consists of 21 open reading frames (Figure 2). Of these, 19 are involved in YP1 biosynthesis. The genes in the Tam cluster have been assigned functions using sequence comparisons to the gene clusters involved in the biosynthesis of prodigiosin, undecylprodigiosin and other NPs (Table 1).[33, 44, 45] 12 of the genes in the Tam cluster have been found to be homologous to the genes involved in the biosynthesis of prodigiosin in *Serratia sp*. (*pig* cluster)[32] and undecylprodigiosin in *Streptomyces coelicolor* (*red* cluster).[31]

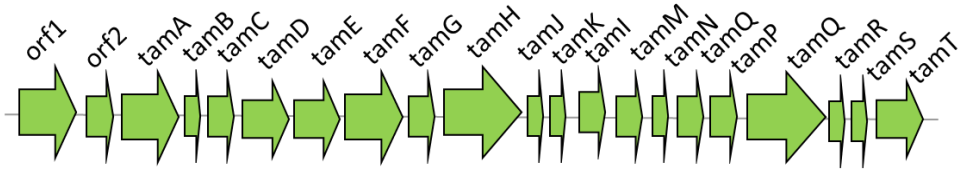


Figure 2: Gene Cluster involved in the biosynthesis of tambjamine YP1. The *tam* cluster consists of 21 open reading frames, 19 encode proteins involved in the biosynthesis of the bipyrrolic natural product.

A bifurcated pathway has also been proposed for the biosynthesis of tambjamine YP1 (Figure 3). Although these steps have not been experimentally verified, the first part of the pathway is similar to the prodiginine biosynthetic pathway proposed by Walsh[46] and Challis[32] in two independent studies. The proposed enzymatic assembly is initiated by the formation of the pyrrole A-ring from L-proline tethered to a peptidyl carrier protein (PCP), TamB via a thioester linkage. The conversion occurs via a four-electron oxidation process by the activity of a flavin adenine dinucleotide (FAD)-dependent dehydrogenase, TamG. The pyrrolyl-2-carboxyl intermediate is then transferred to a ketosynthase (KS), TamF which is responsible for the decarboxylative Claisen condensation of this moiety and a malonyl unit tethered to the ACP domain of TamD. A PLP-dependent decarboxylation step is performed by the α-oxoamine synthase (AOS) domain of TamD in the presence of L-serine, to give the second pyrrole ring. Subsequent oxidation and methylation by a putative dehydrogenase TamJ and a methyl transferase TamP, result in the intermediate product, 4-methoxy-2-2'-bipyrrole-5-carbaldehyde (MBC).[46-48] The final step involves the enzymatic condensation of a primary amine with a long alkyl chain, dodec-3-en-1-amine, derived from the fatty acid biosynthetic pathway and MBC to give tambjamine YP1.

There is a lot of information in the literature about the production of MBC, a key intermediate in the biosynthesis of the prodiginines in *S. coelicolor and Serratia* species.[32, 33, 46, 48] The genetic and biochemical analyses from previous studies have revealed that the genes from the *pig* and *red* clusters that are homologous to the genes in

the *tam* cluster include those likely to be involved in the production of MBC in the biosynthesis of tambjamines in *P. tunicate* (Table 1). Of particular interest to this study are the *pig*H and the *red*N genes, which encodes for a PLP-dependent aminotransferase homologous to *tam*D. Walsh's study of *pig*H has revealed that the protein has three domains, an unusual pair of acyl carrier protein (ACP) domains and a seryltransferase (serT) domain, which is responsible for generating a C2 unit from L-serine used in the formation of pyrrole ring B.[46] Additionally, Challis' study of *red*N gene also revealed that the protein product of this gene carries out the same function as that of PigH protein.[33, 48]

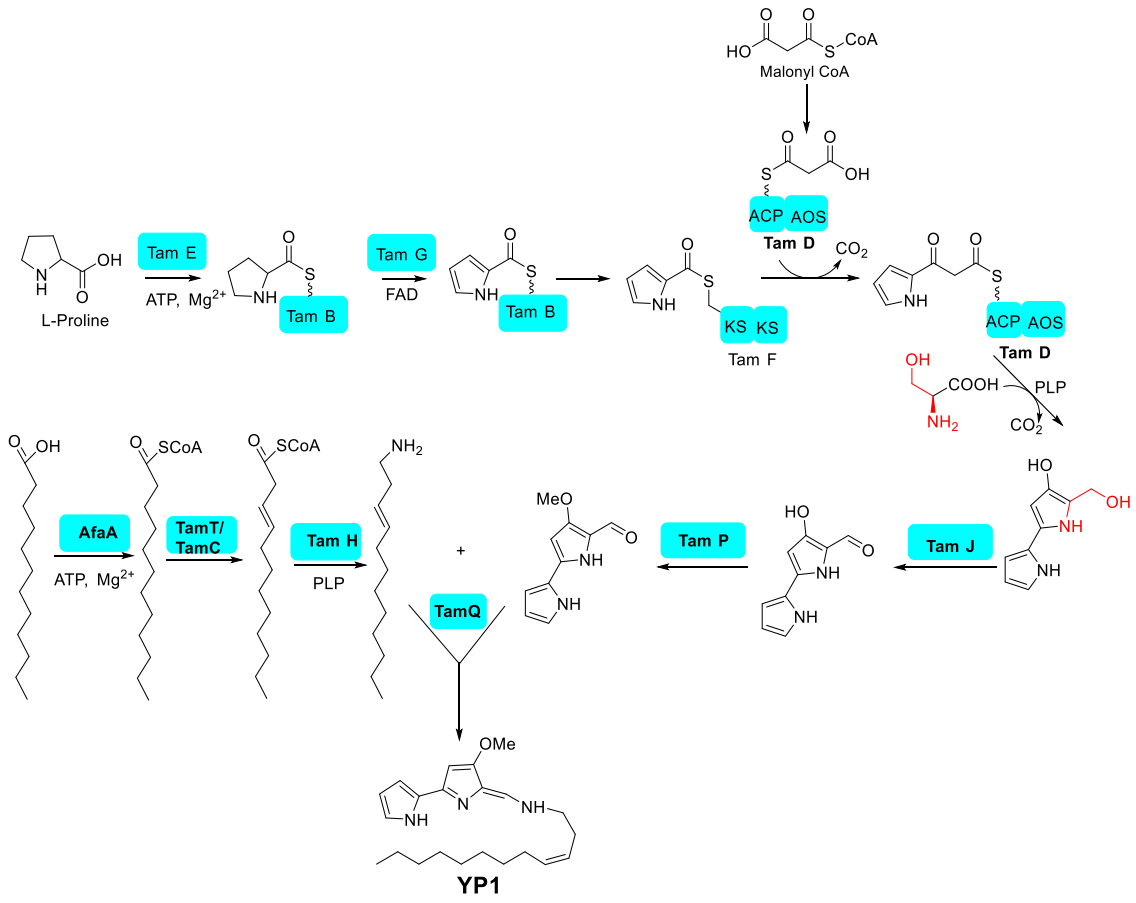


Figure 3: Proposed biosynthetic pathway of the tambjamine YP1 from *P. tunicata*. The bifurcated pathway involves the oxidation of L-proline into a pyrrole ring followed by the condensation of malonyl with L-serine to give a second pyrrole ring. Condensation of the bipyrrolic aldehyde with a long chain amine results in the biosynthesis of tambjamine YP1.

Table 1: Predicted functions of gene products from *tam* Cluster. Each gene product of the *tam* cluster has been assigned a putative function based on sequence homology with protein from the *pig* and *red* gene clusters. Gene sequence was obtained from Genebank for *P. tunicata* D2 (NCBI reference sequence: NZ_AAOH01000002.1)

Genes in Tam Cluster	Predicted Size (kDa)	Pig Cluster Homologs	Red Cluster Homologs	Putative Function
tamA	75	-	-	AMP binding protein
tamB	10	*pigG*	*redO*	Peptidyl carrier protein
tamC	43	-	*redG*	Putative oxidase
tamD	58	*pigH*	*redN*	seryl transferase
tamE	56	*pigI*	*redM*	L-prolyl-AMP ligase
tamF	81	*pigJ*	*redX*	Pyrrolyl-β-ketoacyl ACP synthase
tamG	43	*pigA*	*redW*	L-prolyl-PCP dehydrogenase
tamH	104	*pigE*	-	Putative class III aminotransferase
tamJ	40	*pigM*	*redV*	HBC dehydrogenase
tamK	29	-	-	Hypothetical protein
tamI	50	-	-	Putative permease
tamM	45	-	-	Putative permease
tamN	26	-	-	Putative ABC transporter, ATP binding protein
tamO	45	-	-	Hypothetical protein
tamP	37	*pigF*	*redI*	HBC *O*-methyl transferase
tamQ	100	*pigC*	*redH*	Terminal condensing enzyme
tamR	12	*pigK*	*redY*	Hypothetical protein
tamS	25	*pigL*	*redU*	Phosphopantetheinyl transferase
tamT	62	-	-	Putative dehydrogenase

1.3 The α-Oxoamine Synthases

1.3.1 Role in Natural Products Biosynthesis

The α-oxoamine synthases (AOSs) are an expanding family of enzymes which are at the heart of the biosynthesis of many interesting NPs. These enzymes are involved in the biosynthesis of a diverse selection of primary and secondary metabolites that range from being essential for life to being harmful to life (Figure 4) .[11] These include the signalling

molecules sphingolipids[49], as well as the cofactors biotin[50], vitamin B12, heme, chlorophyll and factor 430[47]. Additionally, these enzymes are involved in the biosynthesis of quorum sensing and anti-fouling molecules in bacteria, which include the bipyrrolic alkaloids, tambjamines produced by marine bacteria and marine invertebrates[40], the tripyrrolic alkaloids prodiginines produced by *Streptomyces, Serratia* and *Actinomycete sp.*[32, 46], the tricyclic perhydropurine alkaloid saxitoxin produced by marine invertebrates[51], fumonisin produced by the filamentous fungus, *Fusarium verticillioides*[52] and the autoinducers cholera autoinducer-1 (CAI-1)[53] and *Legionella* autoinducer-1 (LAI-1)[54] produced by *Vibrio cholerae* and *Legionella pneumophila* respectively. AOSs are also involved in the synthesis and degradation pathways of the amino acids L-tryptophan[47] and L-threonine[55] respectively.

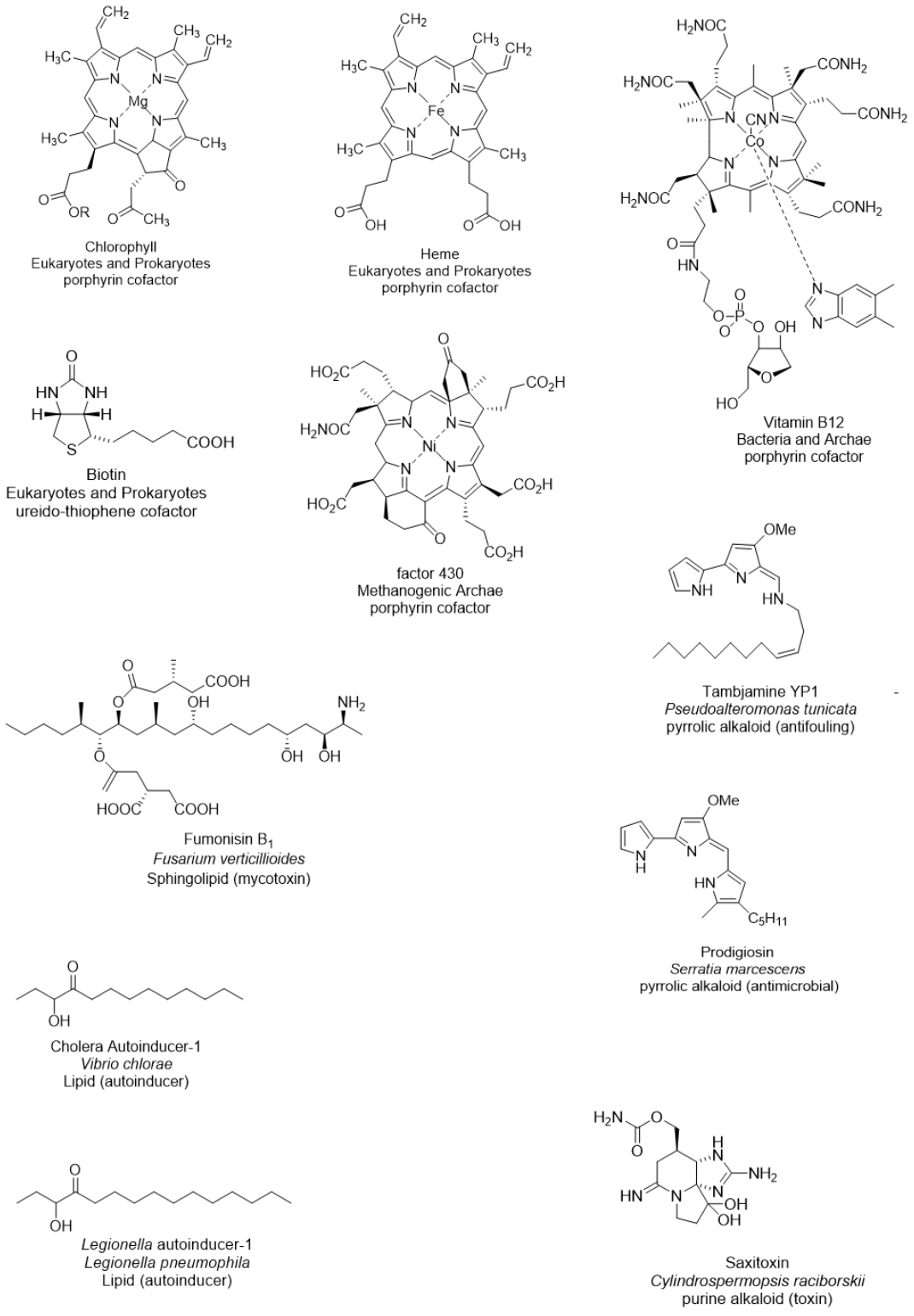


Figure 4: Structures, producers and biological role of molecules from pathways involving AOS enzymes.

1.3.2 Members of the AOS Family

The most well characterised members of the AOS family include serine palmitoyltransferase (SPT)[56-58], 5-aminolevulinate synthase (ALAS)[59], 2-amino-3-ketobutyrate CoA ligase (KBL)[55], 8-amino-7-oxononanoate synthase (AONS)[50] and Cholera CAI-1 synthase (CqsA)[60] (Figure 5 and A1). SPT catalyses the first step of sphingolipid biosynthesis. It also carries out a decarboxylative condensation reaction, and utilizes L-serine and palmitoyl-CoA as substrates to give the first intermediate of the pathway, 3-ketodihydrosphingosine (KDS). This intermediate is then further incorporated into the biosynthesis of ceramides, sphingomyelins, *etc*. SPT is found in mammals, yeast and bacteria and mutation of the human enzyme results in a disorder of the peripheral sensory neurons known as hereditary sensory neuropathy type 1 (HSN1).[56] AONS is a very important enzyme in the biosynthesis of biotin, a cofactor which is required for most carboxylation and transcarboxylation reactions in plants and microorganisms. It catalyses the Claisen condensation between L-alanine and pimelyol-CoA to produce 8-amino-7-oxononanoate (AON).[61] KBL is one of the unique AOS members. It is responsible for catalyzing the second step of the threonine degradation pathway by performing a condensation reaction between glycine and acetyl-CoA to give α-amino-β-ketobutyrate (AKB). It is different from other AOSs in that, it does not perform a decarboxylative step in its reaction mechanism.[55]

ALAS is the first enzyme in the biosynthetic pathway of the porphyrin ring of heme and other tetrapyrroles. It catalyses the decarboxylative condensation between glycine and succinyl-CoA to produce 5-aminolevulinic acid (ALA), the first precursor in tetrapyrrole biosynthesis. ALAS occurs in most prokaryotes and eukaryotes, however mutations in the human erythroid ALAS result in a genetic disorder known as sideroblastic anemia.[59] Interestingly, the gene product of ORF34 from the ECO-02301 biosynthetic gene cluster in *Streptomyces aizunenesis* NRRL-B-11277 which is involved in the formation of the 2-amino-3-hydroxycyclopent-2-enone five membered ring (C_5N scaffold) also utilises the same substrates as ALAS (glycine and succinyl-CoA) to give the condensed product ALA.[62]

CqsA is the most recent member of the AOS family whose structure has been determined in 2010 by Bassler[63], Taylor[53] and colleagues. This enzyme is involved in the biosynthesis of the quorum sensing molecule, 3-aminotridecan-4-one also known as CAI-1 in *V. cholera.* [60] The actual mechanism of the enzyme remains unclear as there are two possible amino acid substrates reported by Bassler's group which the enzyme may catalytically condense with decanoyl-CoA to give CAI-1. These are (*S*)-2-aminobutyrate [60] and (*S*)-adenosylmethionine (SAM)[63]. LqsA is the closest homologue of CqsA with a 42% sequence identity. It was identified in the opportunistic pathogen *L. pneumophila* in 2008 by Spirig *et al*., and is responsible for the synthesis of the α-hydroxyketone signalling molecule LAI-1. It belongs to the *lqs* gene cluster which regulates virulence, biofilm formation and bioluminescence in *L. pneumophila*.[54] More recently JqsA, a homologue of LqsA and CqsA, was identified in *Janthinobacterium sp.* HH01 which is involved in violacein biosynthesis through the production of the autoinducer JAI-1. The chemical structure of JAI-I has not yet been elucidated. However, sequence analysis of JqsA using the NCBI database suggests that the enzyme has properties similar to 7-keto-8-aminopelagonic acid (KAPA) synthase which is involved in the biosynthesis of biotin. This would suggest that JqsA is likely to function as an alanine transferase.[64]

The arginine transferase, SxtA from the saxitoxin gene cluster is also a member of the AOS family. This enzyme is predicted to carry out a condensation reaction characteristic of the family. It utilizes the substrates L-arginine and propionyl-ACP to give the intermediate 4-amino-3-oxo-guanidinoheptane which is further functionalized by the other enzymes in the gene cluster to give saxitoxins.[51] Another interesting member of the AOS family is Fum8p which was identified in the pathogenic fungus *Fusarium verticillioides* by Gerber in 2008. Fum8p is a part of a fumonisin gene cluster that synthesizes the mycotoxin fumonisins. The enzyme is responsible for introducing an amino group and two carbons into the fumonisin backbone and off-loading it in a PLP-dependent fashion from the PKS machinery. This reaction is achieved through a decarboxylative condensation between alanine and stearoyl-ACP.[52]

The AOSs TamD, RedL and PigH are homologues with a seryl transferase function belonging to PKS gene clusters. They are involved in the biosynthesis of the pyrrolic NPs tambjamine, undecylprodiginine and prodigiosin respectively. They accept the amino acid substrate L-serine and carry out a PLP-dependent condensation with a pyrrolyl-*β*-ketoacyl-S-ACP intermediate to give a bipyrrole product, 4-hydroxy-2-2'-bipyrrole-5-methanol (HBM). These enzymes are very different from the other AOSs because they exist as natural fusion proteins with their ACP partners. RedL and PigH have two ACP domains while TamD is predicted to have one ACP domain, based on sequence analysis. Although functional studies have been carried out on RedL and PigH, no structural information is currently available for these enzymes. RedN on the other hand, although a part of the *red* cluster to which RedL belongs, is much more similar to ALAS in terms of sequence homology. It carries out the condensation between glycine and β-ketomyristoyl-ACP to give 4-keto-2-undecylpyrroline.[32]

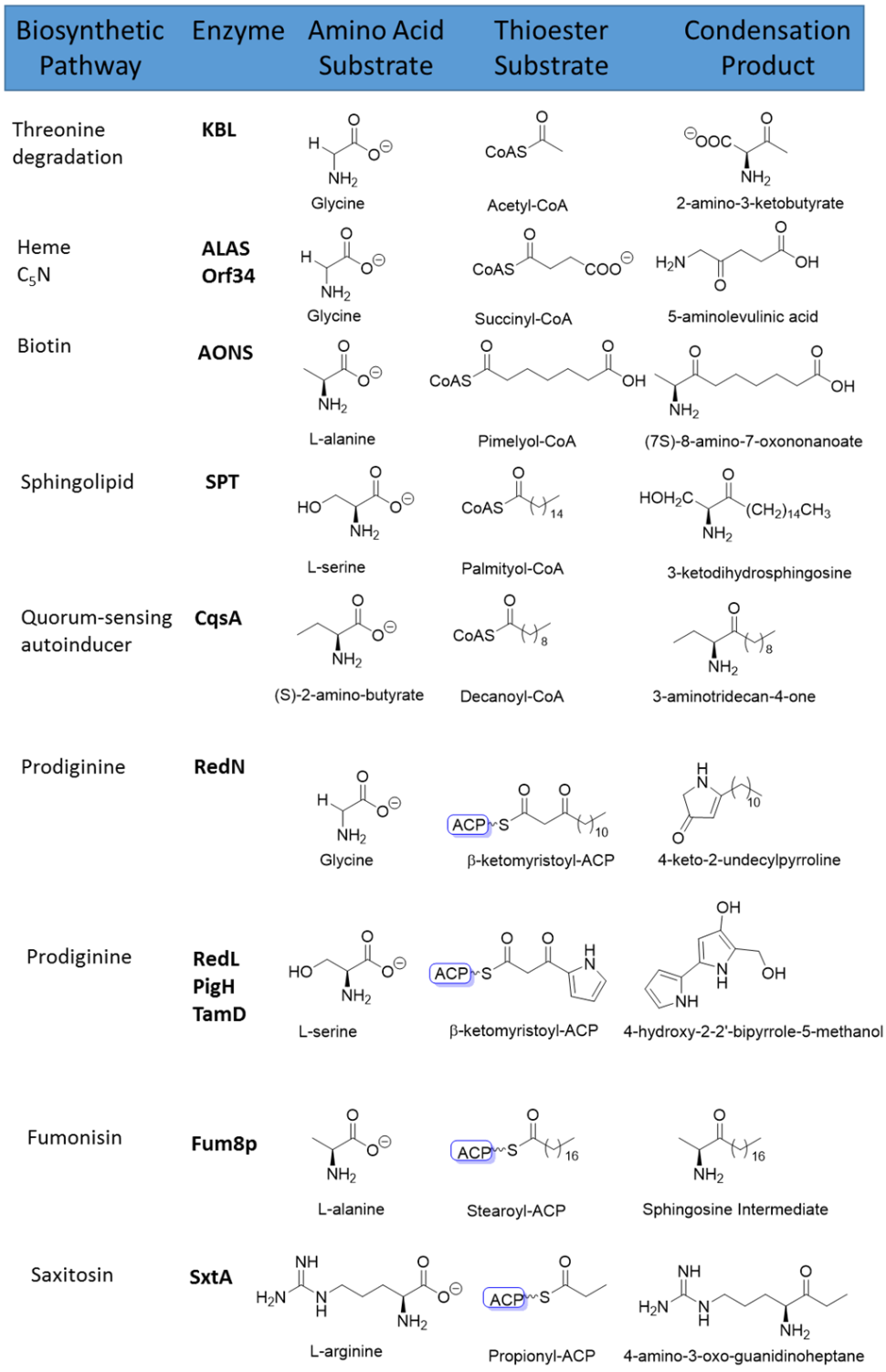


Figure 5: AOS family of enzymes. Each enzyme performs a condensation reaction between an amino acid and a thioester substrate to give an α-oxoamine product.

1.3.3 PLP-Dependent Mechanism

The AOS enzymes catalyse the Claisen-like condensation between their amino acid and acyl-CoA thioester substrates, with a concomitant release of carbon dioxide to give α-oxoamine products (Figure 6).[61] Some members of the family utilise an acyl carrier protein (ACP)-acylthioester instead of the conventional acyl-CoA substrate (Figure 6). So far, KBL is the only member of the family that does not carry out the decarboxylation step to give the final condensation product.[55]

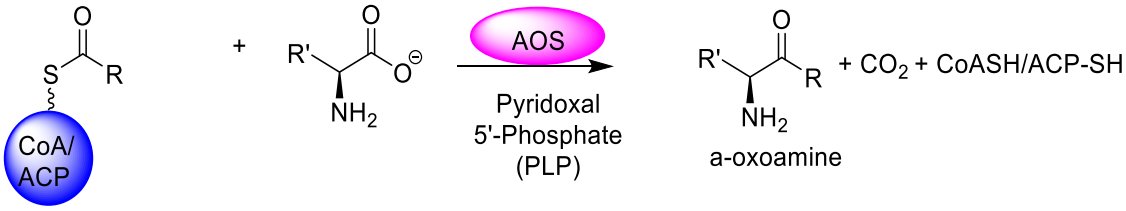


Figure 6: General reaction scheme for AOS enzymes. These PLP-dependent enzymes carry out a Claisen condensation of acyl-CoA thioester and amino acid with concomitant release of carbon dioxide.

All the members of the AOS family discovered to date are dependent on the PLP cofactor for enzymatic activity. These enzymes have a conserved PLP binding site with a lysine residue in the active site forming a Schiff base with the cofactor (internal aldimine).[65] The reaction mechanism utilized by AOS enzymes can be summarized in six steps: (A) displacement of catalytic lysine and binding of amino acid substrate to PLP to form an external aldimine, (B) abstraction of the α-proton to form a resonance-stabilised quinonoid intermediate, (C) nucleophilic attack of the quinonoid intermediate on the acyl-CoA substrate to form a *β*-ketoacid product with concomitant loss of CoA, (D) decarboxylation of the *β*-ketoacid to form a quinonoid product, (E) protonation of the product quinonoid to form an external aldimine and (F) release of the α-oxoamine product from PLP to generate an internal PLP aldimine.[56] These steps are shown below in the reaction mechanism for SPT (Figure 7). In AOS enzymes, the C^{α}-H of the amino

acid-PLP external aldimine is oriented perpendicular to the PLP ring in accordance with the Dunathan hypothesis.[66]

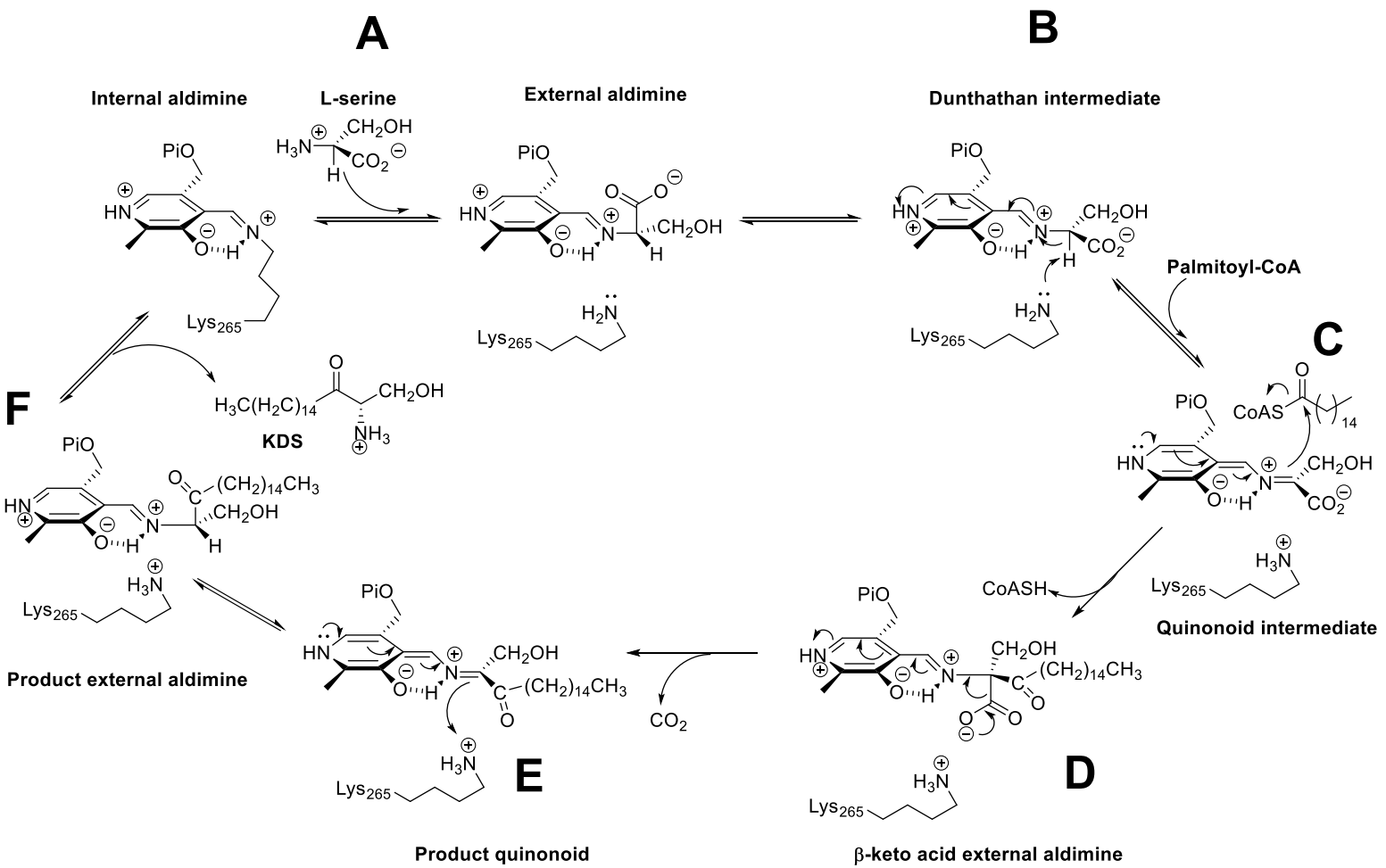


Figure 7: Reaction Mechanism of SPT. PLP serves as the cofactor during the Claisen-like condensation reaction between L-serine and Palmitoyl-CoA in six steps. (A) Formation of internal aldimine, (B) abstraction of α-proton, (C) formation of β-ketoacid, (D) decarboxylation of β-ketoacid, (E) protonation of quinonoid product and (F) release of AOS product.

1.4 Acyl Carrier Proteins

1.4.1 Role in Natural Products Biosynthesis

ACPs are essential for the production of various primary and secondary metabolites. They are small acidic proteins comprising of less than 100 amino acids (~9 kDa), that are involved in providing starter units and 2-carbon chain extension units derived from acyl-CoA needed for molecular assembly of FAs and PKs.[67] Their biosynthetic homologues are known as peptide carrier proteins (PCPs) and are involved in the construction of peptide backbones from amino acid monomers used in NRP biosynthesis.[68] ACPs are also responsible for directing intermediates between the fatty acid synthases (FAS) and the polyketide synthases (PKS), stabilising pathway intermediates and regulating these pathways.[67, 69] Therefore ACPs can be thought of as the "drivers" of NP biosynthesis.

ACPs are expressed in nature in their *apo* form and are functionally activated by the addition of a 4'-phosphopantetheine (4'-PP) prosthetic group derived from CoA, to a conserved serine residue via the hydroxyl side chain to produce the *holo* form of the protein (Figure 8). [70] The post-translational modification (PTM) occurs by the enzymatic activity of an acyl carrier protein synthase (AcpS) or a phosphopantetheinyl transferase (PPTase) in the presence of CoA and magnesium ion (Mg^{2+}).[71] The thiol group of the 4'-PP moiety is responsible for shuttling acyl intermediates of the fatty acid and polyketide biosynthetic pathways among the active sites of enzymes in the assembly line.[72]

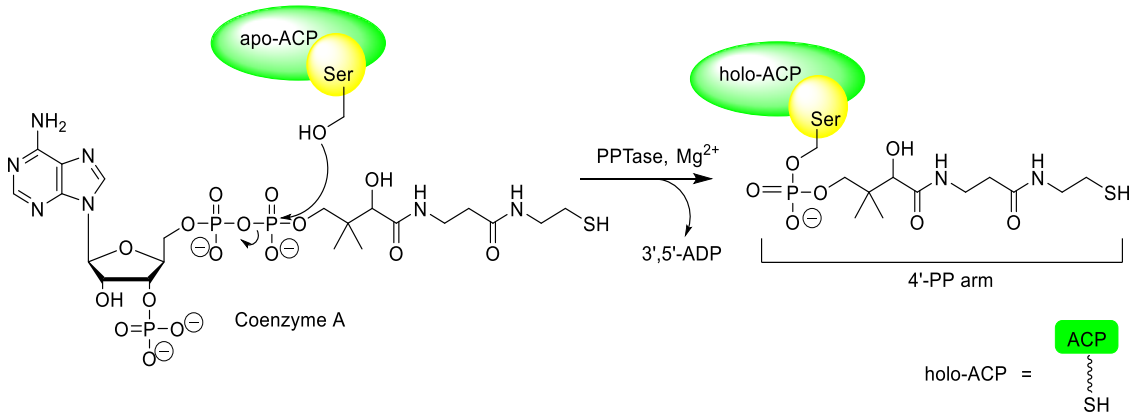


Figure 8: Activation of *apo*-ACP to form functionally *holo*-ACP with 4'-PP prosthetic group. In the presence of a PPTase and Mg^{2+} ions, the phophopantetheinyl arm derived from CoA is added to the ACP via the hydroxyl side chain of serine to form *holo*-ACP.

1.4.2 Type I and II ACPs

ACPs can either exist as part of large multifunctional megasynthase complexes (type I systems) or as discrete proteins (type II systems). Fatty acid and polyketide biosynthetic pathways have both type I and type II ACPs, whereas non-ribosomal peptide synthases (NRPSs) exist mainly as megasynthases with type I carrier proteins.[70] The structural organisation of the type I FAS and PKS are very similar, consisting of up to 12 functional modules per polypeptide chain. For instance in the FAS system each module consists of a ketosynthase (KS), acyltransferase (AT), ketoreductase (KR), enoyl reductase (ER) and dehydratase (DH) domain which play a role during biosynthesis (Figure 9A). Each module may contain a single ACP, as seen in animal FAS or more than one carrier proteins as in type I PKS and NRPS systems, where the growing PK or NRP intermediate is covalently attached via the thioester linkage (Figure 9B-C). [73] Protein-protein interactions between carrier proteins and other catalytic domains involved in FAS, PKS and NRPS pathways are necessary for the formation of intermediates and the final product, as the substrates involved in elongation are anchored by thioester linkages to carrier proteins.[68] For example, a typical PKS elongation cycle

will require the ACP to interact with the AT domain that loads the starter unit, the KS domain which carries out condensation to incorporated the starter unit into the growing chain and the other domains that chemically modify the chain. Likewise, in NRP biosynthesis, the ACP analogue, PCP requires interaction with the adenylation and the condensation domains.[68] In type I FAS, the ACP domain is very flexible and is able to interact with each catalytic domain within the megasynthase in an iterative manner.[74, 75] However, protein-protein interactions between carrier proteins and their enzyme partners in type I PKS and NRPS systems are somewhat restricted. This is because each carrier protein in these systems is responsible for a specific step within the assembly line process which only allows interaction between neighboring domains.[75]

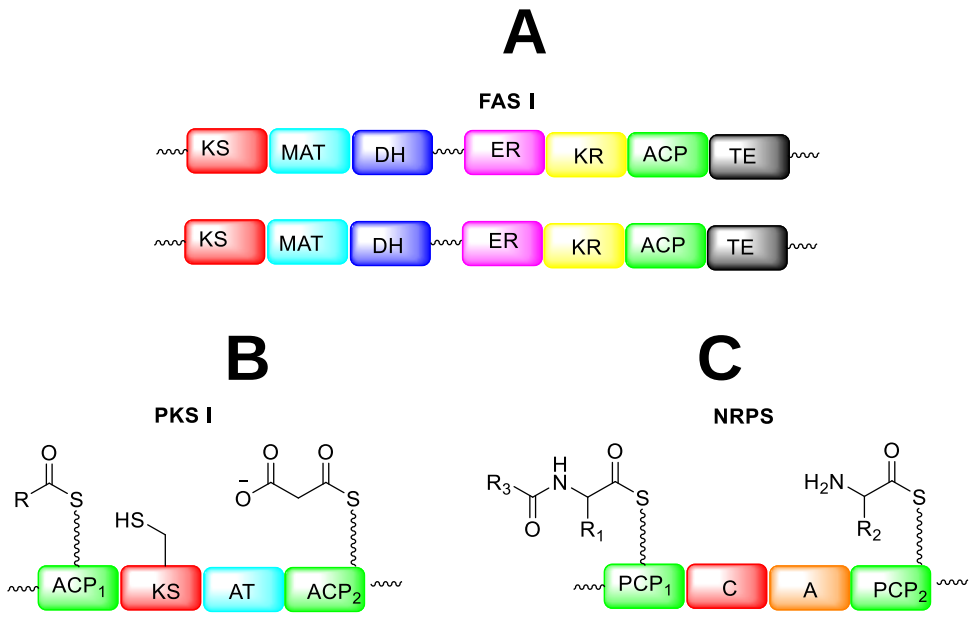


Figure 9: Typical type I domain arrangement for FAS, PKS and NRPS. The carrier protein can interact with each domain in the module. (A) Mammalian FAS consisting of two α-chains. (B) A typical PKS elongation cycle will require the ACP to interact with other domains. (C) In the NRPS system the adenylation (A) domain activates the amnio acid monomer which is incorporated into the growing chain by a PCP. The condensation (C) domain then catalyses amide bond formation between the two monomers tethered to the PCPs.

Type II FAS and PKS systems carry out the same functions as the type I systems. However, in the type II systems, the catalytic domains including the ACPs exist as discrete proteins, *i.e*., separate polypeptides rather than one polypeptide chain. In the type II systems, the growing acyl chain that is covalently bound to the ACPs is not readily accessible because it is usually buried in the hydrophobic pocket of the protein which means that ACP-enzyme interactions are very crucial for the partner enzyme to access its target.

Type I and II ACPs are structurally similar with dynamic and flexible properties. These intrinsic properties are utilized in protein-protein interactions with several enzyme partners in nature. Their structures are greatly stabilized by pantetheinylation and acylation.[75] ACPs possess a conserved structure among many organisms even at low sequence identity.[70] The protein adopts a helical bundle fold, with a characteristic four α-helices architecture, with three of these helices running parallel to one another and the fourth helix (helix III) lying antiparallel to the others (Figure 10).[70, 75] The helices are arranged to form a hydrophobic pocket in the centre of the protein, which houses the growing acyl chain during FA biosynthesis with an opening to the pocket situated between helices II and III.[70] The serine residue which undergoes PTM pantetheinylation is a part of a conserved three amino acid motif (Asp-Ser-Leu, DSL), which is located in the N-terminus region of the second helix.[74] ACPs are very dynamic and highly flexible proteins with helix I being among the most flexible regions of the ACP followed by helices II and III. The dynamic nature is necessary for substrate delivery and interactions with enzymes.[70] The prosthetic 4'-PP arm of the carrier protein allows the protein to span a distance of 20 Å, which also requires additional flexibility in the linker region between domains in the type I systems in order for the protein to effectively shuttle substrates to the active sites e.g. such as between the KS and AT in PKS I systems.[68]

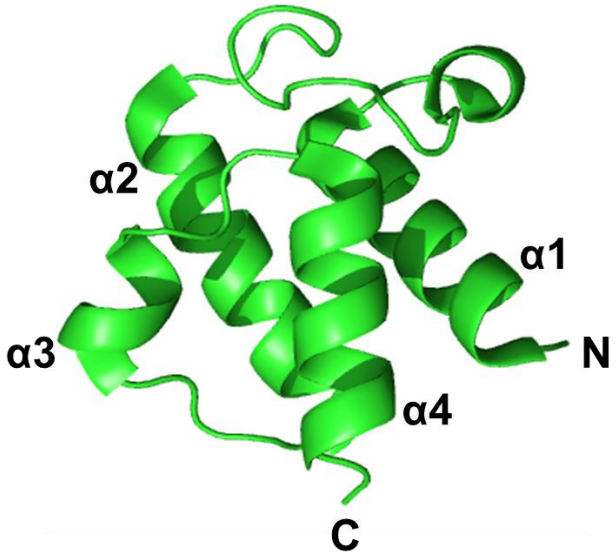


Figure 10: Crystal structure of apo ACP from *E. coli.* ACP adopts a helical bundle fold, with a characteristic four α-helices architecture, with three of these helices running parallel to one another and the fourth helix (helix III) antiparallel to the others, (PBD code: 1T8K).

1.4.3 ACP-Enzyme Complexes

The interactions of ACPs with their enzyme partners are essential for sequestration and processing of acyl substrates during the biosynthesis of NPs and fatty acids. As previously described, the biosynthetic intermediates are usually tethered to the 4'-PP arm of the ACP via a thioester linkage. The ACP delivers the substrate in the correct orientation to the active site of its enzyme partner avoiding exposure to the solvent, preventing the hydrolysis of the delicate thioester bond. This process requires optimal contact for efficient delivery. Structural studies of the yeast FAS has led to the proposition of a chain transfer mechanism described as the 'chain-flipping' mechanism by Cronan. This model suggests that the process begins with the acyl-ACP thioester being buried initially within the hydrophic pocket of the ACP. The entire chain is then flipped into the hydrophobic active site of the respective enzyme partner, where it becomes fully extended (Figure 11).[76]

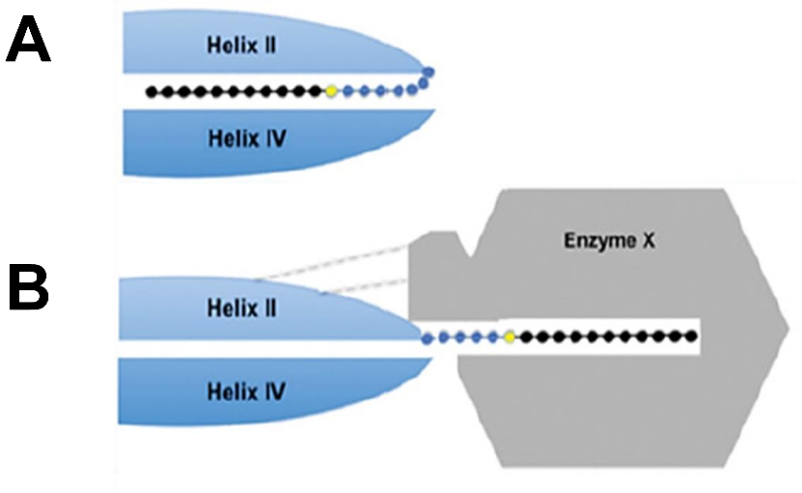


Figure 11: Cartoon depiction of the chain-flipping mechanism of acyl-ACP thioester in complex with enzyme partner.[76] (A) The acylthioester is housed in the hydrophobic pocket of the ACP formed by residues mainly of helices II and IV (the prosthetic group is coloured in blue, the thiol in yellow and the acyl chain in black. (B) The acyl chain and most of the prosthetic group is flipped from the hydrophobic pocket of the ACP into the pocket of the enzyme partner X. (*Taken from reference 73*)

This mechanism has been supported by the crystallographic data from the structural studies of ACP-enzyme complexes. The first ACP-enzyme complex was reported in 2000 by Parris *et al*. from the structural studies of *Bacillus subtilis* ACP with its *holo* AcpS.[77] The structure showed that AcpS interacted largely with helix II of the ACP, which had the attached 4'-PP arm, through mostly hydrophilic interactions in the form of salt bridges between the acidic residues of helix II and arginine residues of AcpS.[77] More recent ACP-enzyme complexes have also been reported. These include acyl-ACP in complex with the following partners; FabA (3-hydroxyl-ACP dehydratase), LpxD (N-acyltransferase), FabI (enoyl reductase), BioH and BioI. The ACP-FabI structure also reveals that protein-protein interactions occurs via the acidic residues of helix II of the ACP (Asp35, Asp38 and Glu41), which form stable electrostatic interactions with basic residues of FabI (Lys201, Arg204 and Lys205). The electron density for the pantetheinyl arm of the ACP was absent from the structure, which hindered the prediction of how ACP delivers its substrate to the active site of FabI (Figure 12A).[78]

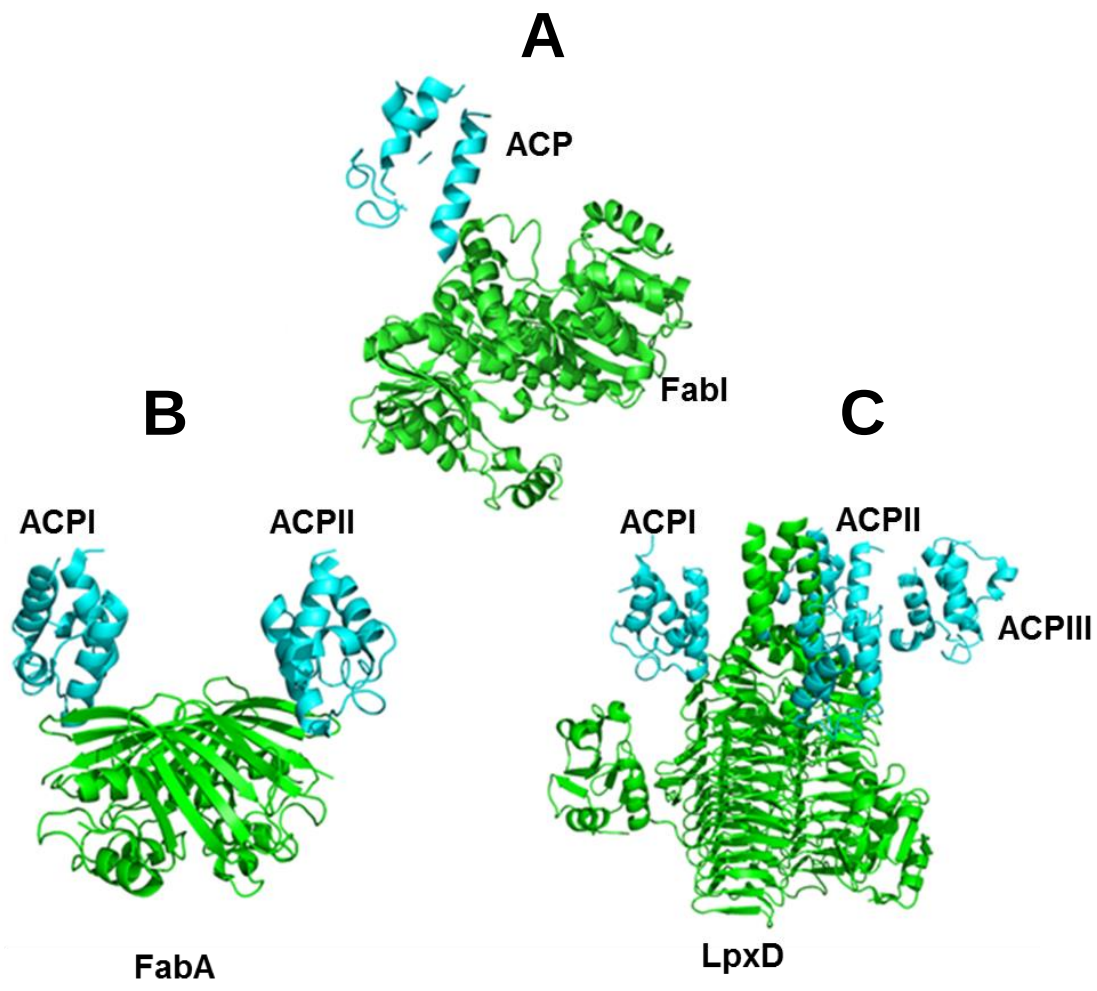


Figure 12: ACP-Enzyme Complexes. (A) *E. coli* ACP in complex with FabI, the FASII enoyl reductase (PDB code: 2FHS) (B) *E. coli* ACP in complex with FabA, 3-hydroxyl-ACP dehydratase (PDB code: 4KEH). (C) *E. coli* ACP in complex with LpxD, an acyltransferase involved in Lipid A biosynthesis (PDB code: 4IHF). ACP coloured in blue and FabI, FabA and LpxD in green.

Nguyen *et al.* later reported the structure of a chemically cross-linked acyl-ACP·FabA complex, in which the acyl chain occupied the hydrophobic cavity formed between the subunits of dimeric FabA (Figure 12B). The structure revealed a collapsed ACP which had a similar topology to that of *apo*-ACP, which suggests a dynamic process involving the ACP moving from a sequestered-substrate state to a 'chain-flipping' state to position the acyl chain within the active site of the enzyme partner.[79] The crystal structure also

showed that the acidic residues of the helices II and III of the ACP that interacted with FabA included Glu41, Glu47, Glu53 and Glu60. The acyl-ACP·FabA crystal structures also captured the ACP in two different conformations. The first ACP·FabA conformation reveals one of the ACPs binding to an arginine-rich groove of FabA and the second conformation shows the ACP completely docked with FabA before delivering its acyl chain. These snapshots reflect the dynamic and transient nature of the ACP-enzyme interactions.[79]

A very interesting crystal structure of the acyl-ACP·LpxD complex was reported by Masoudi which captured the acyl-ACP in three different forms (intact-acyl-ACP, hydrolysed-acyl-ACP and *holo*-ACP) while in complex with LpxD (Figure 12C).[80] These structures showed that the acyl chain had flipped from the hydrophobic pocket of the ACP into a hydrophobic channel formed between to monomers of the trimeric LpxD, where it is fully extended.[80] BioI forms a tight protein-protein complex with acylated-ACP which is processed during FA biosynthesis to produce pimeloyl-ACP. The crystal structure of BioI in complex with ACP acylated with tetradecanoic acid (C14), hexadec-9Z-enoic acid (9-Z-C16) and octadec-9Z-enoic acid (9-Z-C18) revealed the insertion of the entire acyl chain into the active site of BioI upon binding. The phosphopantetheine linker interacts with active site residues of BioI through hydrogen bonding of its carbonyl oxygen and amide nitrogen with the side chain of active site residues. The phosphate group also interacts with BioI residues via water mediated hydrogen bonding. Residues from three of the ACP helices (I, II and III) were involved in the binding of the protein to BioI through salt bridge interactions between acidic residues of the ACP and basic residues of BioI. Other interactions were also observed in the crystal structure between non-polar residues of both the ACP and BioI through hydrogen bonding via the backbone amide groups. It was concluded that the acylated ACP was necessary to stabilize the BioI·ACP complex.[81] The complex of BioH with Me-pimeloyl-ACP revealed a BioH·ACP interface and enzyme-carrier protein interactions which bear similarity to that of the other complexes described previously. The contact of the two proteins at the interface is mediated solely by acidic residues of helix II of the ACP and

the basic residues of helices II and III of BioH. Ionic interactions were observed between Glu34 and Glu37 of the ACP and Arg138 and Arg142 of BioH, and Arg155 and Arg159 of BioH with Glu46 and Asp55 of the ACP. Both sites are located at either end of helix II of the ACP. These interactions help in stabilizing the ACP in the required position to BioH to effectively direct the acyl-thioester chain into the active site.[82]

In summary, the acidic surface residues of helix II play an important role in the dynamic nature of ACPs and their substrate specificity.[75] Additionally, NMR studies have shown that the conserved DSL motif in the amino acid sequence which houses the site of pantetheinylation, located at the beginning of helix II (Figure A3), also assists in protein-protein interaction between ACPs and their enzyme partners. These interactions make it possible for acyl intermediates to be protected by being buried in a hydrophobic core of the ACP until it is transferred to the hydrophobic pocket of the respective enzyme partner. In contrast, type I PKS and NRPS carrier proteins have been observed to have neutral protein surfaces as opposed to a positive surface, which leads to the question of whether a similar ‘chain flipping’ mechanism occurs for PKS and NRPS sytems.

Part II: Antibacterial Properties of Plant Natural Products

1.5 Plant Natural Products

1.5.1 Back to Our Roots

Historically, plants have been the main source of NPs up until the early 1900s. Biodiversity within the plant kingdom has provided us with various classes of NPs which have a broad spectrum of biological activities. Several of these compounds are still widely used medicines today. These include the salicylates from willow bark, morphine from the poppy plant and quinine from cinchona. Several of the drugs approved by the FDA for treating cancer are also plant derivatives, which include vinblastine and vincristine from Madagascar periwinkle (*Vinca rosea)*; taxol from Pacific yew (*Taxus brevifolia*) and camptothecin from the Happy tree (*Camptotheca acuminata*).[83] However, over the years, less focus has been placed on harnessing plant NPs for the discovery of lead compounds due to their increased extinction rates, seasonal and environmental variations which affect their access and supply. Additionally, plant extracts are usually very complex mixtures with the bioactive molecules present only in very small quantities after fractionation. Despite these disadvantages, plants are still a major source of NPs, with 91 plant compounds being included in clinical trials up to 2007.[3] With advance in genomics and metabolic engineering, an alternative approach to using plant NPs is to revisit some of the classical families of compounds that have been previously explored and find new ways of exploiting their biological activities whether as pure compounds, complex extracts or adjuncts to existing synthetic compounds. In this body of work, two families of plant NPs were revisted and applied to separate bacterial protein targets, the chalcone flavonoids and the thiosulfinate, allicin from garlic.

1.5.2 Flavonoids

1.5.2.1 Biosynthesis

The flavonoids are a family of polyphenolic NPs produced by plants via the phenylpropanoid biosynthetic pathway. Flavonoids possess a hydroxylated phenolic backbone, with two aromatic rings (A and B) connected by an α,β-unsaturated system, which can also exist as a pyran ring. Flavonoids are further divided into subfamilies based on variations of the pyran ring which include the chalcones, flavonones, flavones, dihydroflavonols, flavonols, flavans, isoflavonoids, neoflavonoids, anthocyanins and cathechins (Figure 13).[84] Their biological activities are greatly dependent on the structural class, extent of oxidation, degree of conjugation and pattern of substitution around the A and B rings.[85]

Their biosynthesis begins with the deamination of the aromatic amino acid phenylalanine to form cinnamate by the activity of the enzyme phenylalanine ammonia lyase (PAL). This step is followed by hydroxylation at C-4 by cinnamate 4-hydroxylase (C4H) to give 4-coumarate, which is then activated to 4-coumaroyl-CoA by 4-coumarate CoA ligase. The resulting CoA product is then condensed with three acetate units derived from malonyl-CoA by the activity of chalcone synthase (CHS) followed by cyclisation by the same enzyme to give naringenin chalcone. This intermediate is then further functionalized into various flavonoids by other enzymes in the pathway (Figure 14).[85] Flavonoids are synthesized by plants as growth regulators and also as defence agents in response to environmental stress e.g. oxidative stress and microbial infections. They can be found in all parts of plants including the fruit, flower, stem, bark and root as either glycosylates, aglycones or methylated derivatives.[85]

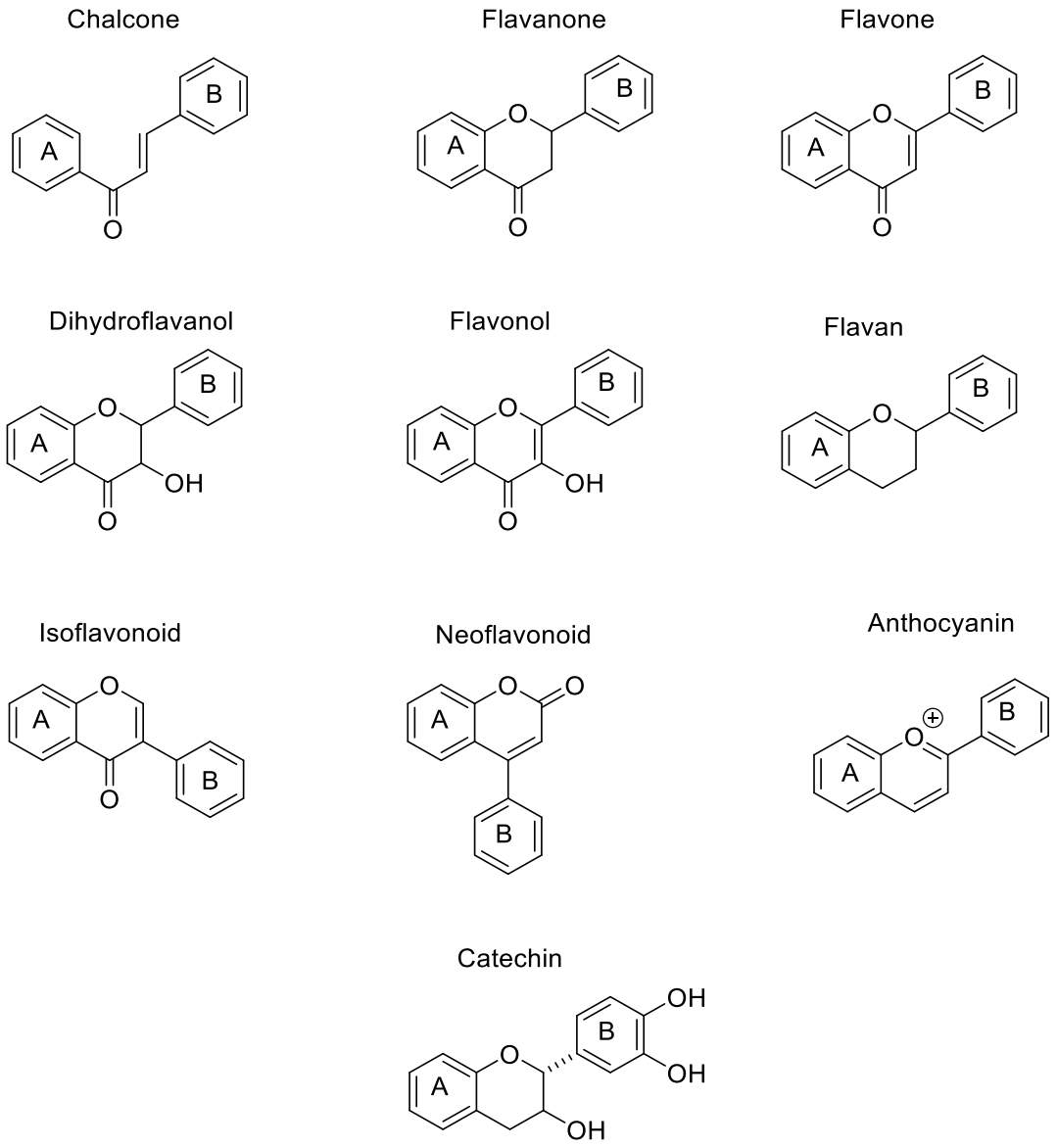


Figure 13: Structural backbone of subclasses of flavonoids. The basic structure of the flavonoids consists of two aromatic rings A and B connected by a pyran ring. The flavonoids also differ based on the extent of oxidation and the presence of substituents on the pyran ring to give nine subgroups.

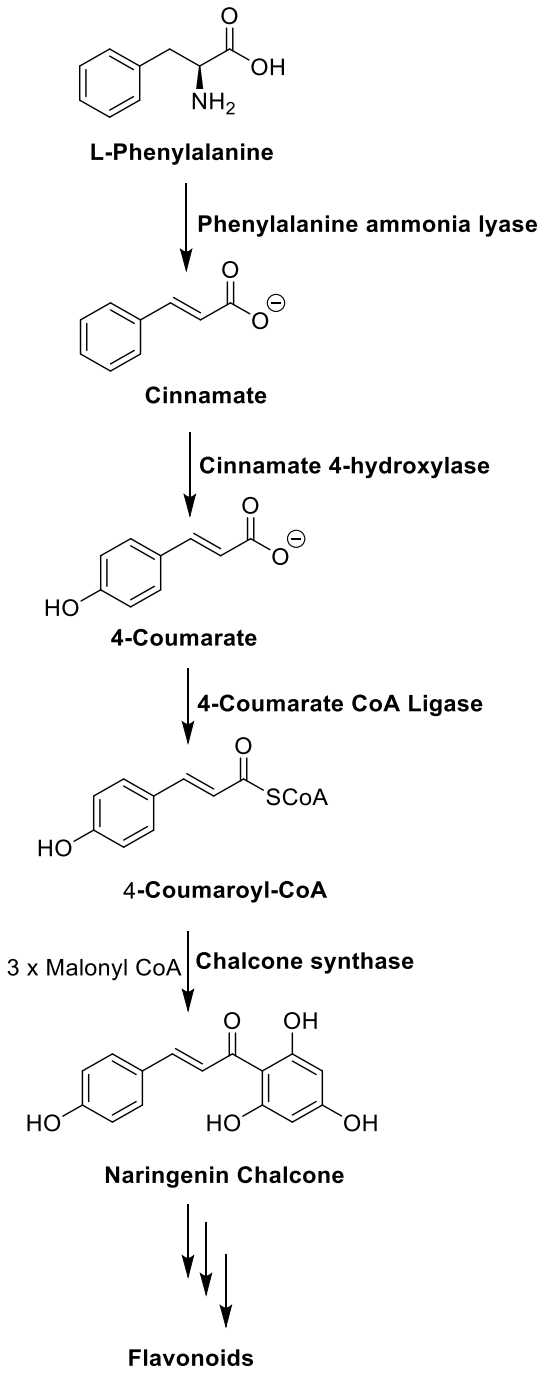


Figure 14: Biosynthesis of flavonoids. Flavonoids are biosynthesized from phenylalanine via the phenylpropanoid pathway. Three units of acetate are incorporated from malonyl-CoA which condenses with the 4-coumaroyl-CoA intermediate to produce naringenin chalcone which can be further functionalized to give structurally diverse flavonoids.

1.5.2.2 Biological Activities

Flavonoids are also known to have a broad spectrum of biological activities including antioxidant, anticancer, anti-inflammatory and antibacterial properties.[84-87] Dietary flavonoids have long been exploited for their associated health benefits, especially for their antioxidant activity. The ability of flavonoids to scavenge radicals and chelate metal ions has been associated with the antioxidant property displayed by these molecules. One such mechanism by which this occurs is through the inhibition of enzymes involved in the generation of reactive oxygen species (ROS) which include microsomal monooxygenase, glutathione S-transferase, mitochondrial succinoxidase and NADH oxidase.[85] Additionally, flavonoids have also been explored for their antimicrobial properties, which not only occur *in vivo* to protect the plant against infections but also *in vitro*. The structural features of flavonoids make it possible for them to interact with several protein targets and inactivate microbial adhesins, enzymes, cell envelop transport proteins and disrupt microbial cell membranes. Catechins have been reported to inactivate cholera toxins in *Vibrio cholerae,* inhibit glycosyltranferases in *Streptococcus mutans* and DNA synthesis in *Proteus vulgaris*. Naringenin, a flavonone present in citrus fruits has been reported to have antibacterial activities against methicillin resistant *Staphylococcus aureus* (MRSA) and *Streptococci*, which is likely associated with the reduction of membrane fluidity in the bacterial cells. Licochalcones A and C, chalcone flavonoids from licorice extracts have also been reported to have antimicrobial activities against *S. aureus* and *Micrococcus luteus*.[84, 85]

1.5.2.3 Potential of Chalcones

Chalcones are one of the most diverse classes of flavonoids, they exist as either intermediates or end products of flavonoid biosynthesis. Structurally, they can be described as 1,3-diaryl-2-propen-1-ones which may exist as a *cis* or a *trans* isomer, with the latter being more thermodynamically stable.[86, 88] It is interesting to note that several chalcones have been isolated from plant sources which have been approved for clinical

trials for the treatment of cancer, viral and cardiovascular disorders. The α,β-unsaturated moiety of the chalcones is regarded as the main pharmacophore and research has shown that there is a loss of bioactivity when this functionality is removed.[86] The most widely studied biological activities of the chalcones include their antioxidant, anti-inflammatory, anticancer and anti-infective properties. As anti-infective agents, chalcones have been shown to have antiviral, antifungal, antiparasitic and antibacterial activities. The screening of chalcones against both Gram positive and Gram negative bacteria has shown that several naturally occurring chalcones possess low MICs (μg/mL) against several clinically relevant species including *S. aureus*, *E. coli*, *Mycobacterium tuberculosis, Bacillus subtilis* and *Pseudomonas aeruginosa*.[88] The chalcone framework serves as an attractive structure that can be further optimized to achieve increased potency useful in the development of novel anti-infective therapies.

1.5.3 Allicin

1.5.3.1 Garlic-Derived Thiosulfinates

Garlic extracts serve as a valuable source of sulfur-containing NPs which have a wide range of biological activities. The chemical basis of the distinctive characteristics of garlic which include its pungent taste and odour, lachrymatory ability and biological effects on humans and microbes have been studied as far back as the 19th century.[89] Cavallito and Bailey were responsible for identifying the main bioactive component of garlic to be allicin in 1944 from analyzing steamed-distilled ethanolic garlic extracts.[89] This then led to the chemical structure and chemical properties of allicin being elucidated over a series of papers by Cavillito and colleagues at The Winthrop Chemical Company.[90-94] Research in allium chemistry took off in later years with notary work in the field carried out by pioneers including Eric Block and Larry D. Lawson whose research careers were focused on studying the chemistry and biology of allium-derived thiosulfinates. Garlic is the most widely studied member of the allium family because of its pharmacological potential which include its anticancer, antithrombotic, anti-inflammatory, antioxidant and antimicrobial properties.[95] Allicin, the main biological

component of garlic, accounts for approximately 75% of the thiosulphinates found in garlic extracts.[96] It is a pale yellow, oily, volatile, and highly unstable liquid with a characteristic garlic odour.

Unlike other secondary metabolites found naturally in plants, thiosulfinates are only produced when plant tissues of alliums are damaged, resulting in the enzymatic activity of alliinase on S-alk(en)nyl-L-cysteine sulfoxides.[95] The dominant sulfoxides present in alliums are alliin, isoalliin and methiin which are found in garlic, onion and chive respectively (Figure 15).[95] Due to the relative instability of thiosulfinates, they are spontaneously converted to linear dialk(en)yl polysulfides e.g. the degradation of allicin to ajoene. Ajoene is the most widely studied bioactive derivative of allicin. Not only is it more stable than allicin, but it also possess pharmacological properties similar to those of allicin.[97, 98] Upon heating, thiosulfinates also degrade to form allyl alcohol and heterocylic polysulfides (Figure 15).[95] The composition of steamed-distilled oils from different species of allium has various sulfides which are characteristic of each member of the allium family.

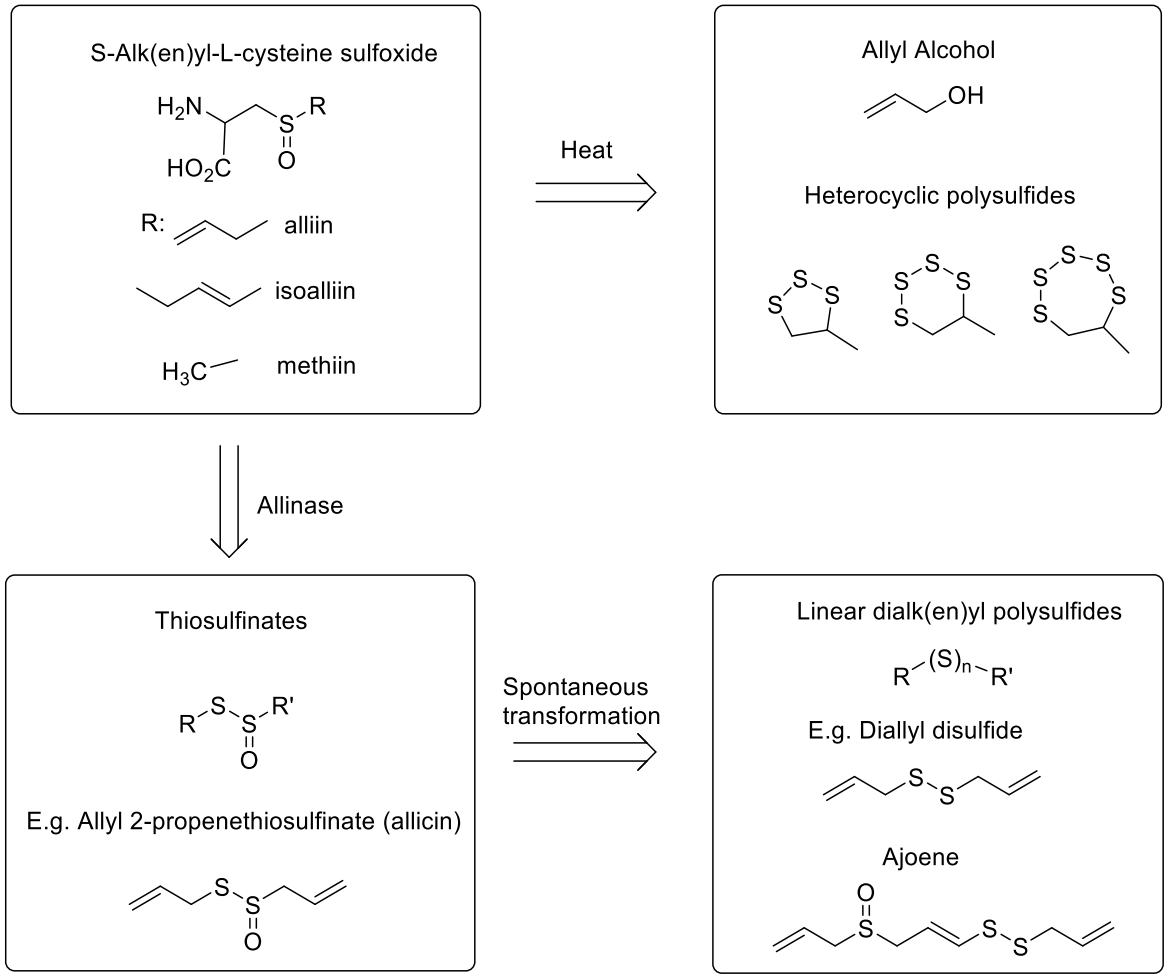


Figure 15: Sulfur-containing compounds in *Alliums* and their transformation products. Alliin, isoallin and methiin are the main sulfoxides naturally occurring in garlic, onion and chive. They can be transformed to other sulfur-containing compounds by heat, enzyme activity or by spontaneous transformation.

1.5.3.2 Biosynthesis

The biosynthesis of allicin in garlic was elucidated by Stoll and Seeback in the 1940s. They showed that the non-protein amino acid, alliin (S-(allyl)-L-cysteine sulfoxide) was transformed into allicin in a two-step mechanism (Figure 16). Firstly, alliin is converted to allyl sulfenic acid, pyruvic acid and ammonia by the PLP-dependent enzyme, allinase when garlic is cut or crushed. The spontaneous condensation of two molecules of allyl

sulfenic acid results in the formation of allicin. This enzymatic reaction can occur within 10-15 minutes at room temperature.

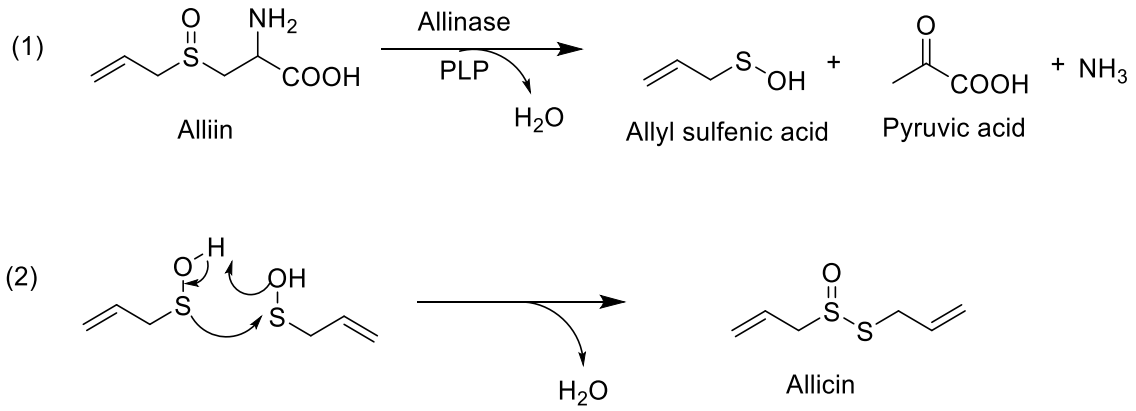


Figure 16: Conversion of alliin to allicin. Two-step mechanism by: (1) which alliin undergoes deamination by allinase to produce allyl sulfenic acid (2) which condenses to form allicin.

Building up on this previous work, Granroth used radioactive labelling experiments to elucidate two possible pathways for the biosynthesis of allicin from serine and glutathione in garlic (Figure 17). His data showed that ^{14}C-labelled serine was incorporated into the formation of ^{14}C-labelled S-allyl-cysteine which would further be used to produce alliin. Additionally, Granroth was able to detect S-allyl-glutathione and S-allyl-γ-glutamyl-cysteine which suggest that gluthathione was involved in the formation of alliin via a second pathway. The source of the allyl group of alliin still remains unclear. On the formation of S-allyl-cysteine, it is further oxidized to form alliin, which then gets enzymatically hydrolysed by allinase to produce sulfenic acid. Spontaneous condensation of two molecules of sulfenic acid gives the product molecule, allicin.[99]

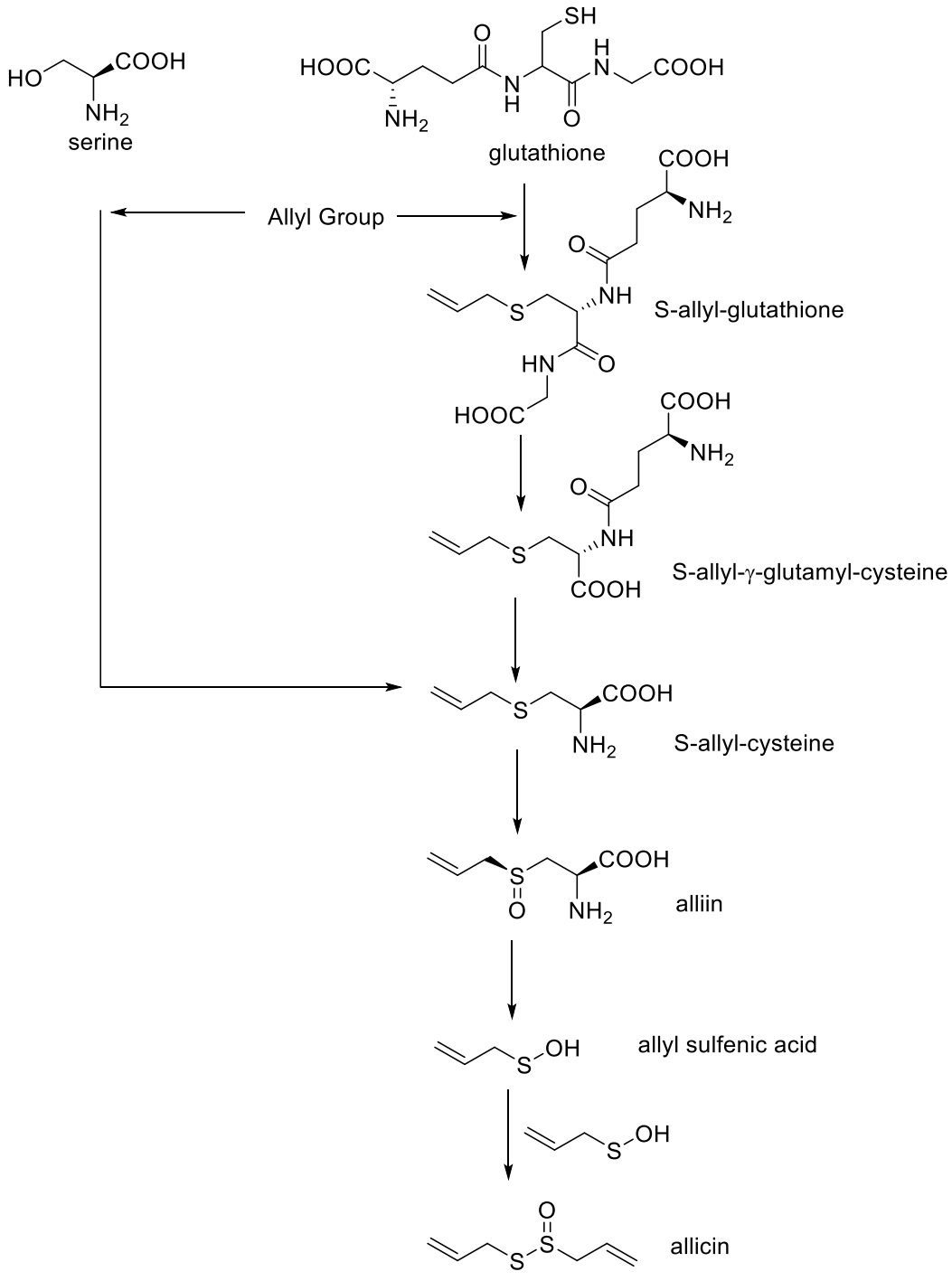


Figure 17: Biosynthesis of allicin. The proposed biosynthetic pathways of allicin start from serine and glutathione. The source of the allyl group is still unclear. Both pathways lead to the formation of the S-allyl-cysteine intermediate. This intermediate is then converted to alliin which then enzymatically forms allicin in a two step mechanism.

1.5.3.3 Antimicrobial Properties

Garlic extracts have been used for thousands of years for their antimicrobial properties against bacteria, fungi, virus and protozoa.[95, 100] Louis Pasteur was one of the first scientists who deliberately used garlic extracts as an antibacterial agent.[99] Garlic extracts have antibacterial activities against several species of Gram positive, Gram negative and acid-fast bacteria (Table 2).[99] Studies have also shown that the allicin present in freshly prepared garlic extracts is the main thiosulfinate that is responsible for the antimicrobial properties of garlic. To support this fact, there has also been good correlation between the percentage concentration of crude garlic and pure allicin extracts used in several investigations aimed at assessing the antimicrobial activities of allicin.[99]

The antimicrobial potential of allicin and aqueous garlic extract (AGE) in comparison with other clinically effective antibiotics has also been assessed in a number of studies. In Fugisawa's study, the anti-staphylococcal potentials of allicin and AGE were observably stronger than that of streptomycin and vancomycin. Additionally, AGE and allicin were found to be more effective against *E. coli* than colistin.[101] Synergism between allicin and other antibiotics has also been reported in the literature.[100, 102-105] In more recent years research has been focused on understanding the mechanism(s) responsible for the antimicrobial activities of allicin and garlic extracts.

In order for allicin to be an effective antibacterial agent, it first needs to be able to reach its target within the cell. Allicin meets this criterion, as it can rapidly diffuse through the bacterial cell wall and the cell membrane. Once inside the cell, it is believed to react with key intracellular thiol-containing proteins required for life, via a disulfide exchange-like reaction (Figure 18).[106] This is further supported by the fact that sulfhydryl compounds such as cysteine and glutathione are able to cancel out the antimicrobial activity of thiosulfinates.[101]

Table 2: Antimicrobial activity of allicin and garlic extracts against Gram positive and Gram negative bacteria. The experimental systems used to assess antimicrobial activity are defined as: complete growth inhibition in liquid culture (A); growth inhibition zones (B); soaked filter discs on top of seeded agar (C); directly pipetted onto bacteria-seeded agar (D); pipetted into wells punched out of the agar (E); LD_{50} (50% lethal dose) determination (F).

Bacteria	Gram	Source of Allicin	Amount of Allicin	Experimental System [A,B,C,D]
Bacillus spp.	+	Pure allicin extract	80 µM[94]	A
		Synthetic allicin	30-150 µM[91]	A
Streptococcus spp.	+	Pure allicin extract	80 µM[94]	A
		Synthetic allicin	60-200 µM[91]	A
MRSA *Staphylococcus aureus*	+	Synthetic Garlic extract	0.0022-0.92 µmol[101]	B,C
MRSA *Staphylococcus aureus*	+	Garlic extract	0.04-0.62 µmol[107]	B,E
Enterococcus faecium	+	Pure allicin extract	>100 µg/mL[100]	F
Salmonella typhimurium	-	Pure allicin extract	80 µM[94]	A
		Synthesized allicin	200-500 µM[108]	A
Agrobacterium tumefaciens	-	Garlic extract	1.72 µmol	B,D
E.coli	-	Garlic extract	0.52-1.72 µmol	B,C
Pseudomonas syringae	-	Garlic extract	1.72 µmol	B,D
V. cholerae	-	Pure allicin extract	80 µM[94]	A
Proteus mirabilis	-	Pure allicin extract	15-30 µg/mL[100]	F
Pseudomonas aeruginosa	-	Pure allicin extract	15 µg/mL[100]	F
Klebsiella pneumonaie	-	Pure allicin extract	8 µg/mL[100]	F
Acinetobacter baumanii	-	Pure allicin extract	15 µg/mL[100]	F

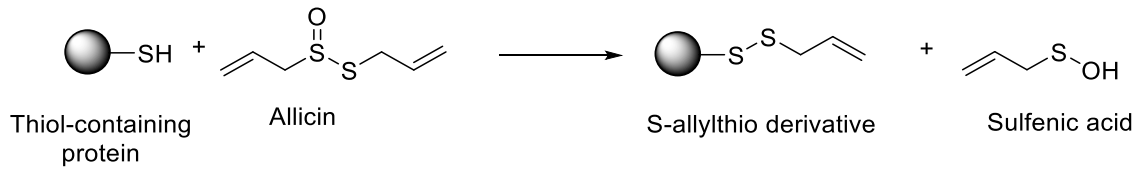


Figure 18: General reaction of allicin with thiol-containing proteins. Free thiols can react with the thiosulfinate moiety of allicin to give S-allylthio derivatives with sulfenic acid as the side product.

A number of important enzymes which are involved in primary metabolic processes from microbial sources have been shown to be inhibited by allicin (Table 3). These include thiol-containing enzymes like cysteine proteinases, the NAD^+-dependent alcohol dehydrogenases, thioredoxin reductases and the non-microbial enzyme papain. Other bacterial enzymes that have been found to be inhibited by allicin includes acetate kinase and phophotransacetyl-CoA synthetase in a non-covalent and reversible manner.[109]

Table 3: Inhibition of microbial enzymes by allicin *in vitro*. Enzymes from four different microbial sources including bacterial and small parasites show inhibition by allicin in a thiol-dependent manner, with the exception of acetate kinase (the bacterial source remains unknown).

Enzyme	Enzyme Source	Allicin Source	Thiol-Dependent Reaction
NAD^+-dependent Alcohol dehydrogenase[110]	*Thermoanaerobium brockii*	Pure allicin	✓
Falcipain 2[111]	*Plasmodium falciparum*	Pure allicin	✓
Rhodesain[111]	*Trypanosoma rhodesiense*	Pure allicin	✓
Acetate kinase/ phosphotransacetylase[109]	Bacterial source	Pure allicin	✘
Cysteine proteinases[112] (Cathepsin B)	*Entomoeba histolytica*	Pure allicin	✓

The antimicrobial activity of allicin has also been believed to arise from inhibition of RNA, DNA and protein synthesis in microbes. Studies carried out by Feldberg *et al.* in 1988 suggests that allicin interferes with RNA, DNA and protein synthesis in

Salmonella typhimurium. Using ^{3}H-labelled uridine, leucine and thymidine, Feldberg was able to show that there was a decrease in the uptake of these building blocks required for RNA, DNA and protein synthesis in the cells treated with allicin.[108] Lipid and fatty acid biosyntheses were suggested to be affected by allicin due to the observed inhibition of acetyl-CoA synthase from different sources.[109]

Part III: Protein Targets of Natural Product Inhibitors

1.6 Sortases

1.6.1 Biological Role

Sortases are cysteine transpeptidases found in Gram positive bacteria. They are responsible for the covalent attachment of a diverse array of proteins on the bacterial cell wall which facilitate interaction with their environment.[113, 114] These interactions include iron acquisition, host invasion and signaling.[113-115] The sortases identified to date have been classified into four sub-families (A-D), based on sequence homology, the cell wall sorting motif of the peptide substrate and their nucleophile substrate (Table 4).[116] Sortase A (SrtA) from *S. aureus* was the first cysteine transpeptidase to be identified in 1999 by Schneewind and colleagues.[117] This class of sortases is also known as the 'housekeeping' enzymes and is highly conserved in Gram positive bacteria. These enzymes are able to recognise protein substrates with an LPXTG motif at the C-terminus and attach them to the pentaglycine crossbridge of the peptidoglycan layer via a lipid II molecule which serves as the second substrate for SrtA enzymes (Figure 19; Figure 21). The protein substrates are expressed in the cytoplasm with an N-terminal signal sequence, a positively charged C-terminal tail and a conserved five amino acid cell wall sorting motif. An active site cysteine residue of SrtA is able to cleave the amide bond between the threonine and glycine residues to generate a thioester intermediate. This acyl-enzyme intermediate undergoes nucleophilic attack by the amino group of the pentaglycine crossbridge of lipid II resulting in the formation of a lipid II-linked product, which is further incorporated into the growing peptidoglycan layer. [114]

Table 4: Classification of sortases. The four families of sortases are based on substrate specificity of the cell wall sorting motif of their protein substrates, cell wall precursor substrate which serves as a nucleophilic sources and biological function.

Family	Protein Substrate	Cell Wall Sorting Motif	Nucleophile	Function
A	Surface proteins	LPXTG	Lipid II	Adhesion, immune evasion, internalization and phage recognition
B	Haem transport factor	NP(Q/K)TN	Peptidogylcan crossbridge	Iron acquisition
C	Pilin proteins	(I/L)(P/A)XTG	Lysine residue of pilins	Pili formation
D	Mother cell and endospore envelope proteins	LPNTA	Lipid II	Spore formation

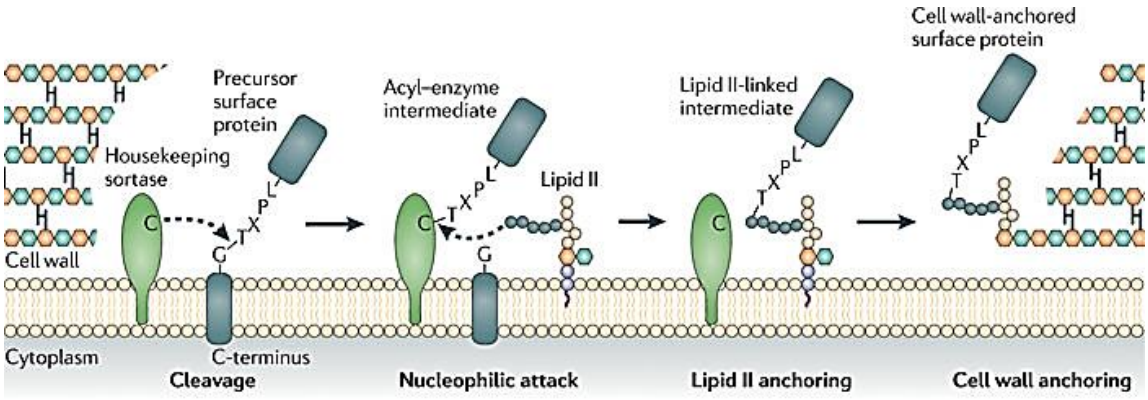


Figure 19: Mechanism of sortase-mediated surface protein anchoring to the cell wall in Gram positive bacteria.[118] The sortase is membrane-anchored and contains an active site cysteine residue which plays a crucial role in the cleavage of the peptide bond between the threonine and glycine residues of the cell wall sorting signal. Nucleophilic attack by a terminal amine on lipid II resolves the thioacyl intermediate to produce a lipid II-anchored protein on the surface of the cell wall. (*Taken from reference 115*)

Sortase B (SrtB) class of enzymes are expressed from the *srtB* gene, which is a part of the iron-regulated surface determinant (*isd*) gene cluster of *Bacilli*, *Clostridia*, *Listeria* and *Staphylococci*. SrtB is only expressed under conditions where iron is limited.[119] This class of enzymes is responsible for anchoring proteins involved in heme-iron uptake that harbours the unique NP(Q/K)TN cell wall sorting motif to the crossbridge amino group of the assembled peptidoglycan. The mechanism is very similar to that of SrtA enzymes.[119] The cross-linked, haem-containing structures are anchored near membrane transporters to carry out their function in iron transport.[120] Sortase C (SrtC) enzymes on the other hand are involved in pili polymerization. The *srtC* gene is only found in spore-forming Gram positive bacteria. The pilin subunits harbour a C-terminal sorting signal similar to that of class A enzymes. SrtC enzymes link together two to three pilin subunits via isopeptide bonds to form the long thin shaft of the pili. The shaft is then attached to the cell wall by class A enzymes which results in a termination of the polymerization reaction. [118] Class D enzymes (SrtD) are the least studied sortases. They are expressed in spore-forming Gram positives e.g. *Bacillus spp*. and *Streptomyces spp*. SrtD enzymes are able to recognize proteins with the LPNTA sorting motif at the C-terminal end and attach the cleaved product to the diaminopimelic acid moiety of the peptidoglycan during spore formation. There is no structural data available on this class of enzymes.[121]

1.6.1.1 Sortase A in Oral *Streptococci*

Sortases play a major role in *Streptococcus* adherence and colonization to its hosts.[122] In a microbiology review on *Streptococci* by Nobbs, it was shown that most of the cell wall-anchored and surface-associated proteins expressed by *Streptococci* bear a C-terminal cell wall sorting motif that can be recognized by sortases.[122] The scope of this study only allows a review of the SrtA enzyme from the oral bacterium, *S. mutans*, which is by no mean a full coverage of the importance of sortases in oral *Streptococci*.

The genome of *S. mutans* reveals the presence of a single sortase enzyme, SrtA which has been demonstrated to be involved in the surface attachment of six proteins expressed by the bacterium with the C-terminal sorting signal (Table 5).[123-127] Studies carried out by Igarashi over a series of four publications between 2003 and 2004 have shown that a deletion or an inactivation of the *srtA* gene results in the inability of *S. mutans* to anchor these proteins to the cell wall resulting in a loss of the cariogenic properties of these proteins e.g. aggregation in the presence of saliva and reduced biofilm formation. Additionally, the Δ*srtA* genotype is not lethal, which suggests that the *srtA* gene is not essential for bacterial growth.[123, 124, 128-130] These findings are very significant to the Oral Microbiology field, as *S. mutans* has been implicated as the main etiological agent of tooth decay in humans. [131]

Another interesting feature of *S. mutans* is that it adopts a biofilm lifestyle in the mouth contributing to the formation of dental plaque, which is an important first step in the formation of tooth caries. The initial stages of biofilm formation require bacterial adherence to oral surfaces which is mediated by surface proteins.[126] Dextranase (Dex) for instance plays a further role in biofilm formation as it not only controls adhesive properties, but also the ability of *S. mutans* to utilize extracellular glucan as a nutrient source.[124, 129] Additionally, the inactivation of *dexA*, the gene responsible for the expression of Dex also results in reduced cariogenicity in a rat model.[132] In addition to these observations in *S. mutans*, mutant strains of *S. gordonii*, (a closely related species) that lacked the *srtA* gene, were observably less likely to colonise oral cavities of mice in comparison to the wild-type strain.[133] These studies therefore suggest that SrtA-mediated interactions in oral *Streptococci* are associated with their pathogenesis.[122] This then makes *S. mutans* SrtA an attractive target for the development of anti-infective therapies.

Table 5: Surface proteins expressed by *S. mutans* and anchored by sortase A enzyme. Six surface proteins expressed by *S. mutans* contains the C-terminal cell wall sorting motif (LPXTG) which is recognized by SrtA. They are all anchored to the cell wall surface and play a role in biofilm formation.

Protein	Abbreviation	Uniprot Code	Function
Glucan-binding protein C	GpbC	Q8DTF1	Dextran-dependent bacterial aggregation
Wall associated protein A	WapA	P11000	Binds collagen for coaggregation
Wall associated protein E	WapE	Q8DU58	Binds collagen for coaggregation
Antigen I/II	Pac/P1/AgB	P11657	Binds collagen, laminin, fibrinogen, salivary and fribronectin
Dextranase	Dex/DexA	Q54443	Hydrolyses glucan and promotes glucan-mediated adherence
Fructanase	FruA	Q03174	Hydrolyses fructose-containing sugars

1.6.2 Structure and Catalysis

The three-dimensional structure of various sortases resolved by NMR spectroscopy and X-ray crystallography revealed that sortases adopt a unique eight-stranded β-barrel core fold, which contains three short helices and random coil loops. This structural arrangement retains high structural conservation among isoforms A, B and C despite class differences (Figure 20). Structural studies of *S. aureus* SrtA also revealed a highly conserved catalytic triad made up of Cys184, His120 and Arg197 (Figure 20 *inset*).[120, 134, 135] Cys184 is the only cysteine residue in the sequence of the protein and is

responsible for the cleavage of the amide bond between the threonine and glycine residues of the LPXTG motif of the protein substrate. This residue is essential for catalysis, as its inhibition by thiol-reactive reagents and mutagenesis renders the enzyme inactive. His120 serves as a general base and Arg197 is believed to stabilise the oxoanion intermediate formed during catalysis.[135, 136] Subsequent mutagenesis experiments of His120 and Arg197 also confirmed that these two residues are essential for the catalytic activity of sortases.[137-139]

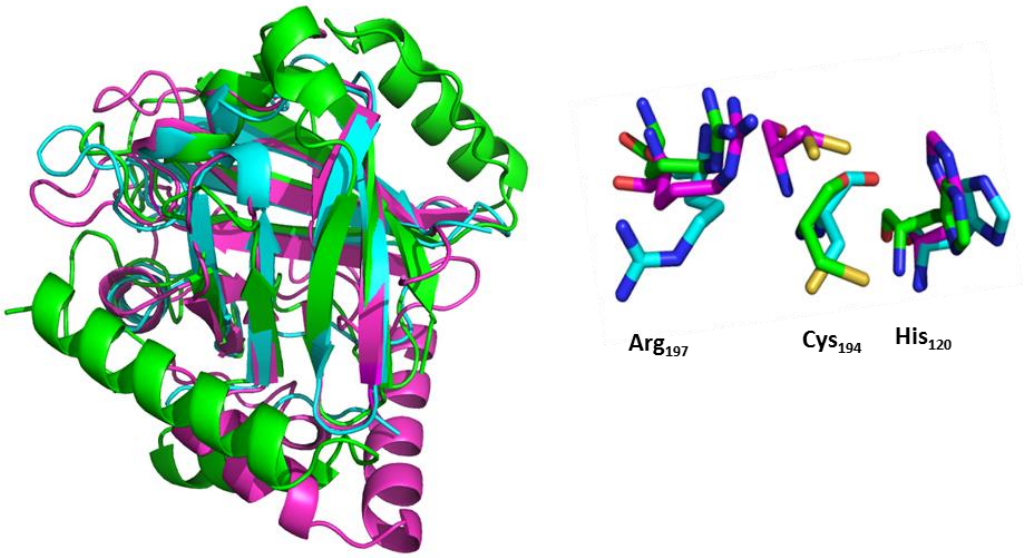


Figure 20: Representative three-dimensional structures of sortase A, B and C. The microorganism, enzyme name and PDB ID are as follows: (Blue) *S. aureus* SrtA, 1T2P; (Green) *S. aureus* SrtB, 1NG5; and (Magenta) *S. agalactiae* SrtC, 3O0P. The Cys-His-Arg catalytic triad is also highly conserved among sortase isoforms (inset).

A general mechanism is proposed for sortases based on the catalytic studies of *S. aureus* SrtA which involves the His-Cys-Arg catalytic triad. The mechanism involves a thioesterification step followed by a transpeptidation step (Figure 21). It has been proposed that the protein substrate for attachment is secreted from the cytoplasm via the Sec secretion pathway and comes in contact with the membrane-bound sortase before anchorage to the cell wall. In the active site of SrtA, a percentage of the enzyme may

have His120 existing in its imidazolium form and the Cys184 in its thiolate form. The LPXTG substrate binds to the active site where it is stabilized by surrounding residues including Agr197. The thiolate form of the cysteine residue then cleaves the amide bond between the threonine and glycine residues of the LPXTG motif via the formation of a tetrahedral thioacyl intermediate stabilized by Arg197. This intermediate then undergoes nucleophilic attack by the terminal amino group of the pentaglycyl side chain of lipid II of the peptidoglycan layer to form a second tetrahedral intermediate which is stabilsed through hydrogen bonding to Arg197. Collapse of the intermediate then results in the formation of a transpeptidation amide product, returning the enzyme to its resting state.[134, 135]

The orientation of His120 and Arg197 in the earlier structures of *S. aureus* SrtA led to inconclusive roles of His120 and Arg197 in catalysis. However, later structural studies of *S. pyogenes* SrtA revealed a novel arrangement of the catalytic residues that led to the accepted roles of His120 as a general base that protonates the leaving group of the tetrahedral intermediate and Arg197 involvement in stabilizing the oxyanion intermediate, substrate binding and substrate orientation through hydrogen bonding.[136] The sortase mechanism was also confirmed for SrtB enzymes in the structural studies of *S. aureus* SrtB by Jacobitz *et al.*, which is very similar to SrtA enzymes.[140]

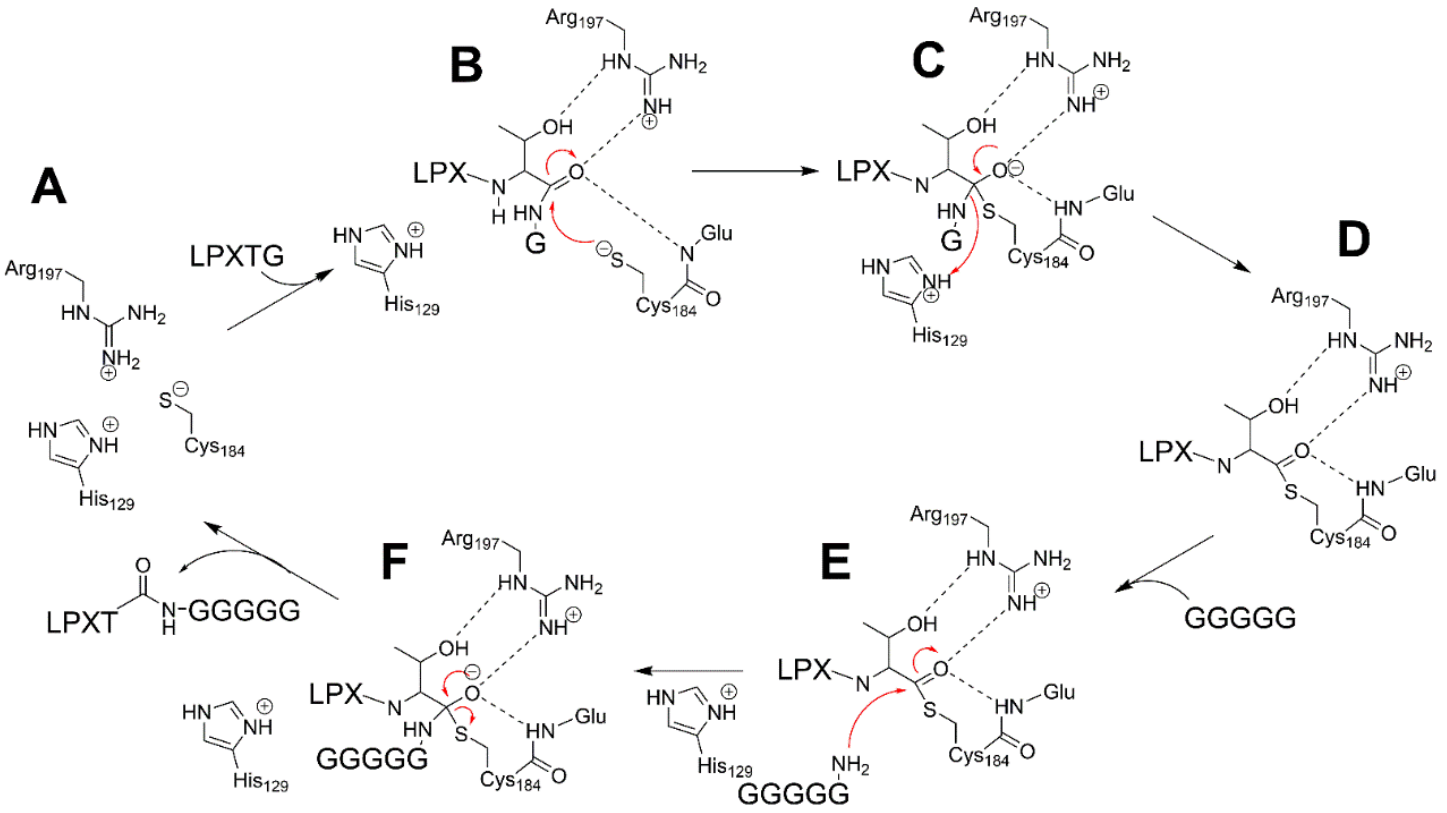


Figure 21: Reaction mechanism of sortase A. (A) Catalytic triad of His-Cys-Arg residues in resting state. (B) Cleavage of the amide bond between threonine and glycine residues by the cysteine residue. (C) Stabilisation of the tetrahedral intermdeiated intermediate to form the thioacyl intermediate D. (E) Nucleophilic attack by the amine group of the pentaglyciine side chain of lipid II on the thioacyl intermediate. (F) Newly formed tetrahedral intermediate stabilized by Arg197 collapses to form the lipid II-linked amide product.

1.6.3 Biological and Chemical Potentials of Sortase A

1.6.3.1 Labelling of Biomolecules

Sortases serve as very useful tools in biotechnology, as they are able to catalyse the site-specific conjugation of diverse array of compounds including small molecules, resins, gels, peptides, sugars, DNA etc. to target proteins.[141] This technique takes advantage of the canonical function of SrtA which involves breaking the peptide bond between the threonine and glycine residues in the primary sequence LPXTG in proteins and making new peptide bonds between the threonine residue and an α-amino group of the pentaglycine side chain of lipid II.[139] Molecules for conjugation are usually functionalised synthetically with oligoglycine tags or LPXTG peptide motifs, making them suitable nucleophiles or SrtA substrates respectively for conjugation to proteins. The range of applications for this technique has expanded over the years, likewise there has been several variations and modifications in the type of substrates used.[141-144] Not only is the sortase-catalysed conjugation method site-specific, but it is also robust and reliable due to the promiscuous nature of the enzyme and ease of purifying soluble, active recombinant enzyme.

1.6.3.2 Target for Anti-Infective Therapy

SrtA has in recent years attracted significant interest as a potential drug target because earlier studies have shown that inactivation of the *srtA* gene had a significant impact upon the pathogenesis of infections caused by several clinically relevant Gram positive pathogens. These include *S. aureus*, *L. monocytogenes*, *S. pyogenes* and *S. pneumonia*.[116] Due to the emergence of antibiotic-resistance strains, infections caused by Gram positive bacteria have posed a challenge for treatment. The attachment of Gram positive bacteria to hosts cells and tissues has been identified as a crucial first step in pathogenesis which is facilitated by sortase-anchored surface proteins. SrtA is therefore considered a universal virulence factor in Gram positive bacteria, and compounds that are able to inhibit enzymatic activity may have potential therapeutic effects against

pathogenic Gram positive bacteria.[113, 116, 145] Several classes of sortase A inhibitors have been identified over the past 20 years since the discovery of these enzymes.[116] These include small molecules from synthetic libraries[138, 146-148], rationally designed peptidomimetics[148, 149] and natural products[150-153]. Recently, Schneewind and colleagues were able to identify a new class of mechanism-based inhibitors, aryl(*β*-amino)ethyl ketones (AEEK) from a screen of over 135 000 compounds which irreversibly modify the active site cysteine residue of SrtA and SrtB.[147] The AEEK inhibitors form covalent adducts with the SrtB via the active site cysteine residue between the *β*-carbon of the inhibitor and the cysteine residue while readily undergoing *β*-elimination (Figure 22).[147]

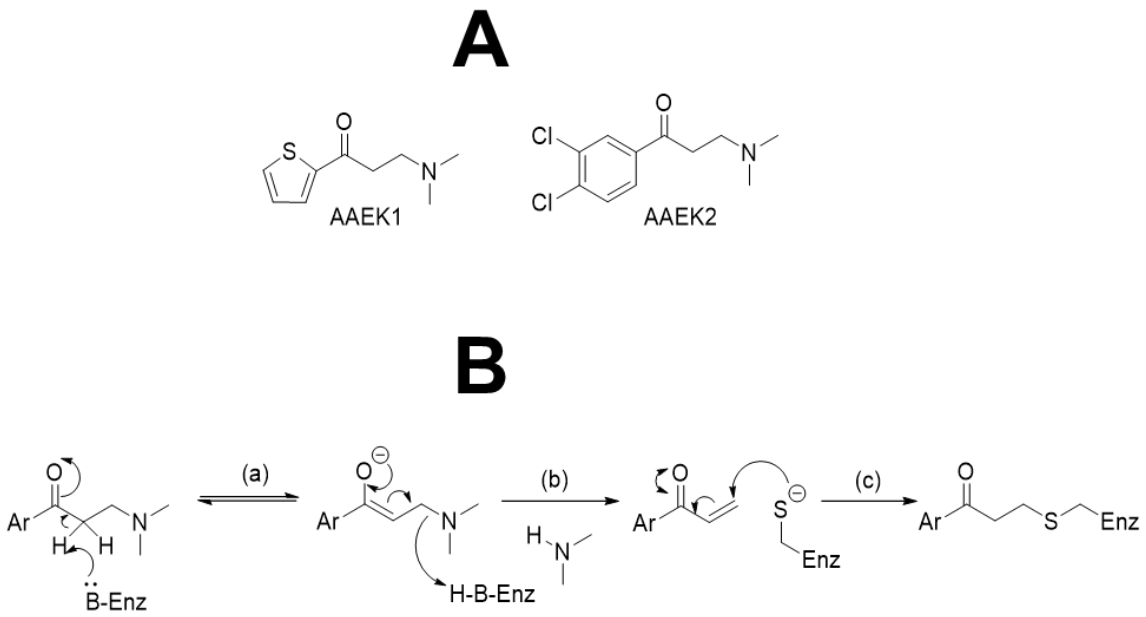


Figure 22: Aryl (*β*-amino)ethyl keotnes inhibitors of sortase A. (A) Molecular structures of the AEEK inhibitors that fromed crystal complexes with *Bacilli* SrtB. (B) Mechanism of sortase inhibition by AEEK inhibitors, (a) active site base of sortase carries out deprotonation of α-carbon, (b) β-elimination of dimethylamine, (c) Michael-type conjugate addition of thiol from the active site cysteine to generate AEEK thioether adduct with sortase.

1.6.3.3 Target for Natural Product Inhibitors

NP screening efforts haved resulted in promising sortase inhibitors from several different classes of NPs (Table 6). The most recent report of NPs with inhibition activity against *S. mutans* SrtA identified two flavonoids, morin[154] and curcumin[145, 155] with IC_{50} values of 27.39 µg/ml and 13.8 µg/ml respectively. Collectively, the results from these studies merit further investigation in identifying NPs that can be further developed for anti-infective therapy. Of equal importance would be further studies aimed at investigating the mechanism of inhibition of sortases by these classes of natural products so as to obtain a better understanding into the kinetics, specificity, affinity and relative effectiveness of the molecules for further drug design.[116]

Table 6: Plant extracts that show inhibition against sortase A *in vitro*. Medicinal plant extracts were tested for anti-sortase activity by monitoring the cleavage of a synthetic fluorogenic peptide. IC_{50} values represent the concentration that reduces enzyme activity to 50% relative to a 1% DMSO control.

Compound	**Source**	**IC_{50} (µg/ml)**
Extract	*Rhus verniciflua*	3.22
Discorhabdins	*Sceptrella sp*.	6.5
Extract	*Fritillaria verticillata*	8.41
Berbine chloride	*Coptis chinensis*	8.7
β-Sitosterol-3-o-glucopyranoside	*Fritillaria verticillata*	18.3
6-hydroxydihydro-β-carboline	*Hyrtios sp*.	67
Palmatine chloride	*Coptis chinensis*	12.7
Kurarinol	*Sophora flavescens*	30.5
Aaptamines	*Aaptos aaptos*	3.7-23.5
Curcuminoids	*Curcuma longa*	13.8-23.8
Cadiolides	*Synoicum sp*.	78.8
B-carboline alkaloids	*Synoicum sp*.	146.4
Sesterterpenes	*Coscinoderma sp.*	36
Bis (indole)alkaloids	*Spongosorites sp.*	15.67
Morin	*Rhus verniciflua*	37.39
Extract	*Achyranthes bidentata*	15.48
Extract	*Benincasa cerifera*	21.57
Extract	*Cibotium barometz*	39.40
Extract	*Cocculus trilobus*	1.52
Extract	*Coptis chinensis*	16.73
Extract	*Cuscuta austrailia*	21.12
Extract	*Evodia officinalis*	13.51
Extract	*Gleditsia japonica*	27.74
Extract	*Liriope platyphylla*	7.96
Extract	*Zanthoxylum bungeanum*	27.29

1.7 Aims and Objectives

This thesis reports data from three independent studies that cover topics introduced in chapters 3 to 5.

Chapter 3 focuses on the structural studies of the putative AOS enzyme, TamD from the tambjamine biosynthetic pathway. In this investigation the objectives were to:

1. Clone, express and purify the TamD enzyme from genomic DNA of *P. tunicata*.
2. Characterise TamD by substrate binding, spectroscopy amd mass spectrometry techniques.
3. Determine the x-ray crystal structure of the enzyme in order to understand the structural relationship and molecular dynamics between the proposed ACP and AOS domains.

Chapter 4 focuses on analysising the molecular basis of SrtA inhibition by *trans*-chalcone. The objectives of this study were to:

1. Clone and purify SrtA from *S. mutans* with an N-terminal truncation of the first 40 amino acids in order to obtain soluble and active protein.
2. Characterise the enzyme using a FRET-based assay which monitors the cleavage of a fluorogenic peptide substrate by the enzyme.
3. Analyse the inhibitory properties of *trans*-chalcone against *S. mutans* SrtA *in vitro* and investigate the mechanism of inhibition using a combination of mass spectrometry, enzyme kinetics and x-ray crystallography.
4. Determine the structure of *S. mutans* SrtA and in complex with *trans*-chalcone.
5. Investigate the impact of chalcone inhibitors on biofilm formation by *S. mutans*.

Chapter 5 focuses on studying the antimicrobial properties of garlic-derived allicin against the multi-resistance pathogen Bcc. The objectives of this study were to:

1. Quantify the amount of allicin present in the freshly prepared garlic extracts and determine its detection limit from the analysis of aqueous solutions of allicin by HPLC.

2. Determine the MICs and MBCs of aqueous garlic extract (AGE) and aqueous allicin standard (AAS) and correlate this activity to the concentration of garlic-derived allicin.
3. Determine the efficacy of various assays including broth/agar dilution, paper disc and agar well diffusion in analyzing the antimicrobial potential of AGE and AAS against Bcc.
4. Use high resolution mass spectrometry to determine the reaction mechanisms of ASS and AGE with a recombinant peroxiredoxin protein (BCP) from *B. cenocepacia.*

Chapter 2: Materials and Methods

The three projects used common techniques, methods and reagents. Where appropriate, the target-specific methods for allicin, TamD and SrtA have been outlined.

2.1 Materials and Reagents

All reagents, chemicals and media were purchased from Sigma-Aldrich, Fisher or Bio-Rad. All primers were purchased from Sigma Genosys and competent cells and pET plasmids from Novagen and Invitrogen. Chromatography columns were purchased from GE Healthcare. HPLC column was purchased from Phenomenex. The *P.tunicata* genomic DNA was purchased from DSMZ, Germany. Synthetic gene for *S. mutans* recombinant sortase A was purchased from GenScript, USA. Allicin was purchased from LKT Laboratories, Germany and bacterial sortase II substrate was purchased from Cambridge Biosciences, UK.

2.1.1 Competent *E. coli* Cell Lines

Table 7: Bacterial strains and applications

Cell line	Genotype	Application
BL21(DE3) (Novagen)	F– *ompT hsdS*(r_B^- m_B^-) *gal dcm (DE3)*	Transformation/ Expression host.
DH5α (Invitrogen)	F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1	Transformation host for plasmid storage.

2.1.2 Growth Media

Growth media and LB-agar plates were prepared by dissolving the components in distilled water (Table 8), adjusting the pH to 7.5 and sterilising by autoclaving for 20 mins at 120 ^oC. They were then stored at 4 ^oC and used within 1 week of preparation.

Table 8: Components used for preparing growth media solutions

Media	Components
Luria Bertani (LB)	Tryptone (10 g/L), yeast extract (5 g/L), sodium chloride (10 g/L)
SOC	Tryptone (20 g/L), yeast extract (5 g/L), sodium chloride (10 mM), potassium chloride (2.5 mM), magnesium chloride (10 mM), magnesium sulphate (10 mM), glucose (2% w/v)
LB-Agar	Tryptone (10 g/L), yeast extract (5 g/L), sodium chloride (5 g/L), agar (15 g/L)

2.1.3 Solutions and Buffers

Buffers were prepared by dissolving the required components from stock solutions in distilled water according to Table 9. All buffers were passed through a 0.2 μm filter and degassed before using for chromatography. Buffers were stored at 4 ^{o}C and used with two week of preparation. The stock solutions were prepared as follows:

1 M tris(hydroxymethyl)aminomethane (Tris-HCl), pH 8.0

5 M Sodium Chloride

1M Imidazole

0.5 M Ethylenediaminetetraacetic acid (EDTA)

25 mM PLP

0.5 M Potassium Phosphate (KPhos) pH 8.0

1M 2-(N-morpholino)ethanesulfonic acid (MES.NaOH) pH 6.5.

Table 9: Components for buffer preparation

Buffer	Components
A	25 mM Tris-HCl, 500 mM NaCl, 125 μM PLP, 10% glycerol, pH 8
B	50 mM Tris-HCl, 500 mM NaCl, 50 μM PLP, 10% glycerol, 1 mM DTT, pH 8
C	20 mM MES.NaOH, 20 % glycerol, pH 6.2
D	20 mM MES.NaOH, 2M NaCl, 20 % glycerol, pH 6.2
E	20 mM MES.NaOH, pH 6.5, 150 mM NaCl
F	50 mM Tris-HCl, 500 mM NaCl, 20 mM imidazole, pH 7.5
G	25 mM KPhos, 500 mM NaCl, 1 mM DTT, pH 8
H	50 mM Tris-HCl, 500 mM NaCl, 250μM PLP, 1 mM DTT, pH 8
I	25 mM Tris-HCl, 1mM EDTA, pH 8.5
J	50 mM NH_4Ac, pH7.5
K	20 mM MES.NaOH, pH 6.5, 125 mM NaCl, 5 mM NH_2-Gly_3 and 2 mM DTT
L	50 mM CAPS, 50 mM MES, 100 mM Tris
M	50 mM acetic acid, 50 mM MEs, 100 mM Tris

2.1.4 Antibiotics

All stocks were prepared by dissolving the required amount of antibiotic in distilled water and then filter sterilised by passage through a 0.22μm filter. Antibiotic was added to cool media to the desired final concentration based on Table 10.

Table 10: Preparation of antibiotics solutions

Antibiotic	Protocol
Ampicillin	100 mg/mL stock solution was prepared and stored at 4 ^oC. The required volume was added to media to give a final concentration of 100 μg/mL
Kanomycin	30 mg/mL stock solution was prepared and stored at 4 ^oC. The required volume was added to media to give a final concentration of 30 μg/mL

2.2 DNA Amplification

2.2.1 Plasmids

The following plasmids outlined in Table 11 were used for cloning and protein expression.

Table 11: Plasmids and their application

Plasmid	Purpose	Application
pGEM-T Easy (Promegra)	Cloning in *E. coli*	Compatible with blue/white screening to identify transformants with ligated DNA product. Provides large amount of stock DNA for manipulation into expression vector.
pET-22b	Expression in *E. coli*	Protein can be expressed with a C-terminal hexa-histidine tag.
pET-28a	Expression in *E. coli*	Protein can be expressed with an N-terminal hexa-histidine tag.

2.2.2 Primers

Primers used in polymerase chain reaction (PCR) for amplification were designed to be between 20-30 base pairs in length (Table 12). Primer oligonucleotides were purchased from Sigma Aldrich and stored at -20 °C as 10 μM stocks.

Table 12: Primers used for molecular cloning and PCR. Bases in bold lettering are the sites of mutagenesis.

Primer	Sequence (5'-3')
TamD forward	CATATGACAGACAATAAAAACACCG
TamD reverse	CTCGAGTGCCACAAAACCGGGTAATTTCA
H139A forward	GCTAGC**GCT**CATGTTTTTGGTATGACCG
H139A reverse	AAAACATG**AGC**GCTAGCAAGAGCATAA
G94A forward	GTAATT**GCC**GGAATTGCCATTCCAGACTTA
G94A reverse	CAATTCC**GGC**AATTACAGGAAGTTTCTGA
SmSrtAN89H139A forward	CATATGAAACTTCCTGTAATTGGCGGAATTGCCATTCC
SmSrtAN89 H139A reverse	CTCGAGTTAAAATGATATTTGATTATAGGAC

2.2.3 Polymerase Chain Reaction (PCR)

The reaction mixture was prepared on ice by mixing forward primer (1.5 μL), reverse primer (1.5 μL), dNTP (1 μL), reaction buffer (5 μL, 10 ×), DNA template (2.5 μL), Pfu Hotstart DNA polymerase (1 μL, 2.5 U) and filter sterilised water (37.5 μL) to obtain a final volume of 50 μL. The thermal cycling protocol was then carried out in a PCR machine according to the steps in Table 13. For A-tailing reaction, the 50 μl reaction from above was added to PCR tube with one Taq Ready-To-Go PCR beads and incubated for 20 mins at 72 ^{o}C. The reaction is then terminated by adding DNA loading dye (2 μl, x5) and the product purified by electrophoresis, which is then cloned into pGEM T-Easy vector.

Table 13: PCR protocol for sequencing

Step	Cycle	Temperature/ ºC	Time/min
Initial Denaturing	1	95	0.5
Denaturing	12	95	0.5
Annealing		54	1
Extension		68	1min/kbp
Final Extension	1	68	10

2.2.4 Site Directed Mutagenesis (SDM)

SDM reactions were performed using primers with overlap extension[156] and the thermostable Turbo Pfu polymerase which exhibits a high fidelity 3'-5' proof reading ability. Each reaction was assembled as follows: Pfu 10x buffer (5 µl), dNTP mix (1 µl), turbo cloned Pfu polymerase (1 µl), forward and reverse primer (2.5 µl) and template DNA (2 µl). The final reaction volume was made up to 50 µl with sterile distilled water and placed in the thermocylcer based on the programme in Table 13.

2.2.5 *Dpn*I Digest

*Dpn*I digest was performed on the SDM reaction on completion of PCR cyle by adding *Dpn*I mix (1 µl) and *Dpn*I buffer (5 µl) to the reaction mixture and incubating for 1 hour at 37 ^{O}C to destroy template DNA. The newly synthesized DNA with mutation (4 µL) was transformed into C2987 competent cells (50 µL).

2.3 DNA Purification and Analysis

2.3.1 Gel Electrophoresis

Agarose (1 g) was added to TAE buffer (100 mL, x1), the mixture was heated until the agarose dissolved. Upon cooling to approximately 50 ºC, gelred (Cambridge Bioscience) (10 µL) was added to the solution which was mixed thoroughly by swirling. The molten

agarose was then poured into a casting mould and allow to set at room temperature. The gel was then transferred to an electrophoresis tank containing TAE buffer (x1). DNA samples containing DNA loading dye (final dilution of x1) was added to gel along with the appropriate DNA marker in separate lanes. Samples were run at a constant voltage of 100 volts for 50-60 mins.

2.3.2 Gel Extraction

The DNA of interest was cut from the agarose gel then weighed and purified using the gel extraction QIAquick kit (QIAGEN). In summary, three volume (3 times the weight of the gel slice) of QG buffer was added to the gel slice and the sample incubated for 50 ^{o}C until the agarose dissolved. One volume of isopropanol was then added to the sample which was transferred to a spin column and centrifuged. The supernatant was discarded and the column washed with 500 µL of QG buffer followed by 750 µL of PE buffer. The residual buffer was removed by centrifuging for 1 minute and the purified DNA was then eluted with 20 µL of EB buffer and stored at -20 ^{o}C.

2.3.3 Ligation

The amplified gene obtained from purification step (3 µL) was incubated with plasmid (1 µL), T4 ligase (1 µL, 1 x 10^6 ligations/unit), T4 ligase buffer (1 µL, x10) and dH_2O (4 µL) for 2 hours at room temperature. The ligation product (4 µL) was then used to transform C2987 competent cells (50 µL). DNA cloned into pGEM T-Easy vector were grown on S-Gal agar plates supplemented with ampicillin (100 µg/mL) for blue/white screening in order to identify viable clones.

2.3.4 Restriction Digest

The following reaction mixture was assembled for a single analytical restriction digest. DNA (8 µL), restriction buffer (1 µL, x 10, New England Biolabs), restriction endonuclease (1 µL, New England Biolabs). BSA (1 µL, x 10) was added to the reaction mixture for enzymes that required it for activity. Reaction components were scaled up for large scale reaction. The components were mixed and then incubated at 37 ^{o}C for 2 hours and the reaction terminated by adding DNA loading buffer. The digestion products were loaded onto an agarose gel for sepration by electrophoresis.

2.3.5 Mini-prep

Plasmid DNA was purified from overnight cultures by following the steps outlined in the manufacturer's protocol which is summarised as follows. Overnight culture grown in LB media (5 mL) was centrifuged at 3000 rpm for 10 mins at 4 ^{o}C. The supernatant was discarded, the pelleted cells resuspended in buffer P1 (250 µL) and then transferred to a sterile eppendorf. The mixture was vortexed until it was homogeneous, after which buffer P2 (250 µL) was added and the sample thoroughly mixed by inverting the tube 6 times. Buffer N3 was then added to the sample followed by another thorough mixing. The resulting mixture was centrifuged for 10 mins at 13 000 rpm and the supernatant decanted into a QIAprep spin column. The sample mixture was then centrifuged for an additional minute, the flow through was then discarded and the spin column washed with buffer PB (500 µL). The spin column was further washed with buffer PE (750 µL), centrifuged for 1 minute and the flow through discarded. The residual wash buffer was removed by centrifuging for an additional 1 minute. The spin column was then placed in a clean 1.5 mL microcentrifuge tube and the plasmid DNA eluted with buffer EB (50 µL) after centrifuging for 1 min.

2.3.6 Sequencing

A PCR was carried out to verify the DNA sequence using thermostable turbo Pfu polymerase. The sample for sequencing was prepared by mixing the DNA template (5 µL), pET or pGEM primer (1 µL, 10 µM stock) (Table 14), sequencing buffer (2 µL) and Big Dye master mix version 3.1 (2 µL). The sequencing reactions were done in pairs with the forward primer used for one reaction and the reverse primer for the second reaction. The sequencing programme consisted of 25 cycles with the following steps: 95 ^{o}C for 30 seconds, 50 ^{o}C for 20 seconds and 60 ^{o}C for 4 minutes. Samples were then analysed by GenePool at the Ashworth Building, The University of Edinburgh.

Table 14: Primers used for sequencing DNA in pET and pGEM vectors

Primer	Plasmid	Sequence (5'-3')
pET forward	pET	TTAATACGACTCACTATAGGG
pET reverse	pET	CTAGTTATTGCTCAGCGGT
pGEM T7 forward	pGEM	TAATACGACTCACTATAGGG
pGEM T7 reverse	pGEM	ATTTAGGTGACACTATAGAA

2.4 Protein Expression

2.4.1 Transformation

Plasmid DNA (2 µl) was added to competent *E. coli* cells (25 µl) and incubated on ice for 20 minutes followed by heat shocking cells for 40 seconds at 42 ^{o}C. The cells were then placed on ice for a further 2 minutes. SOC media (100 µl) was added to the cells which were then incubated at 37 ^{o}C while shaking at 200 rpm for 1 hour. Using sterile techniques, 30 µl of the culture was evenly spread on LB Agar plates supplemented with appropriate antibiotic. Plates were then incubated overnight at 37 ^{o}C.

2.4.2 Single Expression

A single colony was picked into LB broth (500 ml) supplemented with ampicillin (100 μg/mL) or kanomycin (30 μg/mL), and grown overnight at 37 ^{o}C. The overnight culture was used to inoculate fresh LB media (1 L) and cells were grown to an OD_{600} of approximately 0.6 at 37 ^{o}C before induction with IPTG to a final concentration of 0.1 mM. The cultures were incubated for a further 4 hours at 20 ^{o}C after which they were harvested by centrifugation at 5000 × g for 15 minutes. The supernatant was discarded and the cell pellet was stored at -80 ^{o}C until needed for purification.

2.4.3 Cell Lysis by Sonication

The cell pellet was resuspended in lysis buffer with one EDTA-free protease inhibitor cocktail tablet (Roche). The cell suspension was sonicated on ice 15 min with 30s intervals. The resulting crude lysate was centrifuged at 18 000 rpm for 20 min at 4 ^{o}C.

2.5 Protein Purification

2.5.1 Protein Purification for Chapter 3

2.5.1.1 Nickel Resin Purification

The supernatant after the centrifugation step in section 2.4.3 was incubated on equilibrated Ni-NTA resin with an addition of 10 mM imidazole for 1 hour at 4 ^{o}C. The resin was washed with buffer A containing 50 mM imidazole and the protein eluted with buffer A containing 300 mM imidazole. The protein was dialysed overnight into buffer B.

2.5.1.2 Tev Protease Cleavage

The pEHISTEV construct and the Tev Protease used for this step were gifted from the Naismith's group at Scottish Structural Proteomics Facility, University of St. Andrews.

TamD was expressed from a pEHISTEV construct (Figure 23), which consists of an N-terminal hexa-histidine tag followed by a Tev protease site. An N-terminal cleavage at the Tev Protease site was performed following the protocol of Liu and Naismith.[157] In summary, the purified protein after dialysis in buffer B was incubated with His-tagged TEV protease (1 mg/mL of protease for 10 mg of protein) for 16 h at 4 ^{O}C with mild agitation. The digest reaction mixture was incubated with equilibrated Ni-NTA beads for 45 mins and the cleaved protein eluted with buffer B containing 500 mM imidazole. The protein was then extensively dialysed overnight in buffer B to remove excess imidazole.

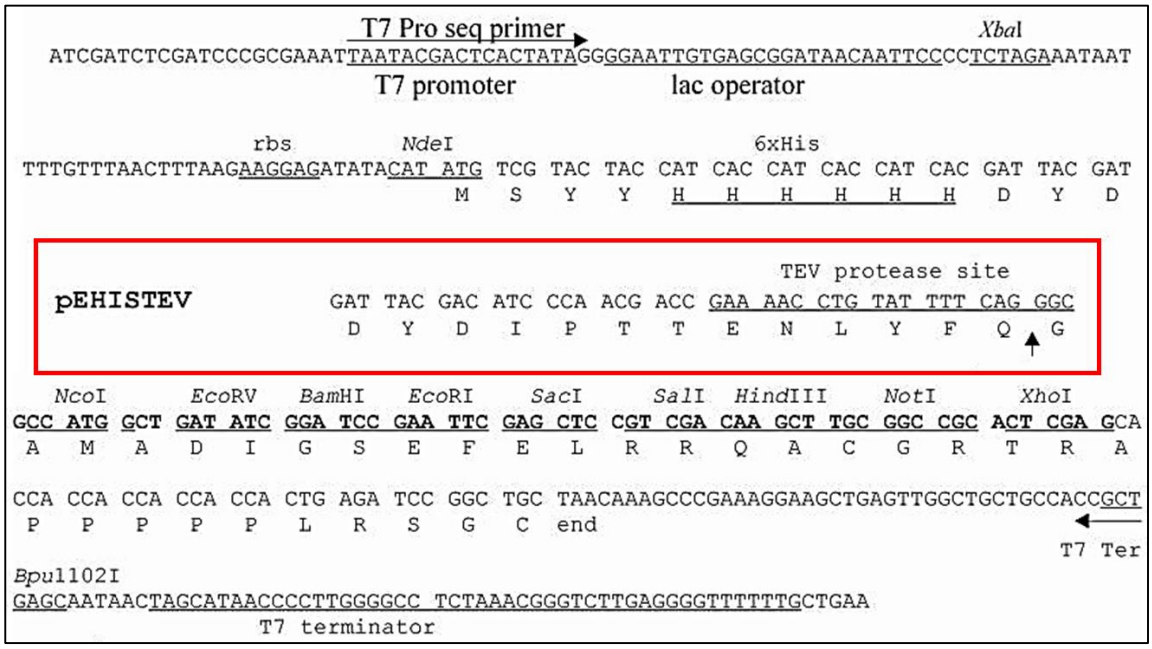


Figure 23: pEHISTEV Plasmid Map. Sequence details of the cloning and expression region of the pEHISTEV vector showing the hexa-histidine and TEV protease recognition sites highlighted in red.[157]

2.5.1.3 Size Exclusion Chromatography (SEC)

The Superdex S200 16/60 column (GE Healthcare) was equilibrated with 1.5 column volumes (CV) of buffer B. The concentrated protein sample (5 mL) was injected on to the column and eluted at 1 mL/min monitoring with UV detection at 280 nm. The

approximate molecular mass of the purified protein was calculated using a calibration curve of known molecular masses (Figure A2).

2.5.2 Protein Purification for Chapter 4

2.5.2.1 Cation Exchange Chromatography

The cell free extract was prepared as previously described in section 2.4.3. Then it was loaded onto a MonoS 5/50 cation exchange column (GE Healthcare) equilibrated with buffer C at 1 ml/min. This was followed by a wash of the column with 15-20 mL of buffer C. The protein was then eluted with an increasing concentration of buffer D over 20 column volumes.

2.5.2.2 Size Exclusion Chromatography

The fractions containing the protein of interest from the cation exchange purification step were pooled and concentrated to 5 mL. The concentrated protein was further purified by SEC using a Superdex S200 column 16/60 (GE Healthcare) equilibrated with buffer E.

2.5.3 Protein Purification for Chapter 5

2.5.3.1 Purification of BCP

A cell free extract was prepared from the cell suspension in buffer F as described in section 2.4.3. The resulting supernatant was filtered (0.45 μM) and loaded onto a HisTrap FF 1 mL column pre-equilbrated in buffer F. The His-tagged BCP protein was eluted from the nickel column with increasing imidazole concentration in buffer F. The sample was dialysed for 16 h in buffer F to remove the imidazole.[158]

2.6 Protein Analysis

2.6.1 SDS-PAGE

Protein samples were denatured by the addition of 2 × SDS sample buffer and boiling for 5 mins. The samples (10 µl) were loaded onto the gel along with low molecular weight (LMW) marker (GE Health Care) which was used for detecting the approximate molecular weights of the proteins. Gels were run in 1× TGS running buffer at 200 V and 180 mA for 50 mins. They were then stained with coomassie blue stain at 37 ^{o}C while shaking gently for an hour and destained by microwaving for 15 mins in water. The acrylamide gel consisted of two layers, running and stacking gels (Table 15).

Table 15: Preparation of acrylamide gels for SDS-PAGE analysis

Component	15% running gel	4% stacking gel
dH_2O	5.7 ml	2.9 ml
40% acrylamide	6 ml	0.75 ml
1.5 M Tris (pH 8.8)	4 ml	-
0.5 M Tris (pH 6.8)	-	1.25 ml
SDS (10% w/v)	150 µl	50 µl
APS (50 mg/ml)	300 µl	100 µl
TEMED	20 µl	5 µl

2.6.2 Concentration using Absorbance, A_{280}

Protein concentrations were determined by measuring the UV absorption at 280 nm using a Varian's Cary 50 UV-Vis spectrophotometer. Calculations were done using the Beer-Lambert's law, (Equation 1). The extinction coefficient (ε) for each protein was determined from the VectorNTI protein analysis software. Spectrum measurements were taken between 200-800 nm.

$$A=\varepsilon cl \qquad \textit{(Equation 1)}$$

(A) = absorbance at 280 nm; (ε) = molar extinction coefficient ($M^{-1}cm^{-1}$); (c) = protein concentration (M), (l) = the path length of the sample (1 cm).

2.6.3 Spectroscopic Measurement

This section is specific for TamD analysis. Excess PLP was removed from all protein samples by passing the concentrated purified sample through a PD_{10} (Sephadex G-25M) desalting column (GE Healthcare) and eluting with buffer G before UV/vis analysis. All UV-visible measurements were done using 500 µl of sample at 37 ^{o}C unless otherwise stated. A Varian's Cary 50 UV-Vis spectrophotometer was used for measuring the UV-visible spectra of all samples and the Cary WinUV software (Varian) used for analysis of results. Samples were analysed with quartz cuvettes with a pathlength of 1 cm from 200-700 nm.

2.6.3.1 Dissociation Constant, K_d

The dissociation constant of TamD was determined in buffer G (38 µM) made up with L-serine (0-100 mM) to a final volume of 400 µl. The reactants were mixed and allowed to equilibrate for 15 mins at room temperature. The absorbance maximum at 426 nm was measured at different concentrations of L-serine and the values plotted against substrate concentration. The data points were fitted to hyperbolic saturation curve (Equation 2) using Sigma Plot software:

$$\Delta A_{obs} = \frac{\Delta A_{max}\,[\text{L-serine}]}{K_d + [\text{L-serine}]} \qquad \textit{(Equation 2)}$$

ΔA_{obs} is the observed change in absorbance at 426 nm and ΔA_{max} is the maximum absorbance change.

2.7 Protein Chemistry

2.7.1 Reactions for Chapter 3 (TamD Analysis)

All reactions for this section are specific for TamD unless otherwise stated. Samples were prepared for the detection of the Lys residue to which PLP forms an internal aldimine with the AOS domain of TamD by reduction with $NaBH_4$ and DTT followed by LysC digest as described in the following sub-sections.

2.7.1.1 Reduction with $NaBH_4$ and DTT

TamD (5 mL, 4 mg/mL) was dialysed for 1 h into buffer H and excess PLP removed using PD-10 desalting column. The protein (450 µL, 11.23 µg) was then mixed with $NaBH_4$ (50 µL) to give a final concentration of 1 mM and incubated on ice. The reaction was monitored by UV/vis spectroscopy until there were no more visible spectral changes at 426 nm. The reduced protein was then denatured with DTT using the following assembly: TamD (5 µL, 11.23 µg), dH_2O (25 µL), buffer I (×5, 10 µL) and DTT (200 mM, 2 µL). The reaction mixture was incubated for 45 mins at 60 ^{O}C and then allowed to coold to room temperature before adjusting the volume to 50 µL with dH_2O.

2.7.1.2 LysC Digest

The protein digest with Lys-C endoproteinase was performed according the protocol outline by the supplier Roche. In summary, lyophilised Lys-C was reconstituted with 50 µL of dH_2O to give concentration of 0.1 µg/µL. The Lys-C proteinase (0.5 µg, 5µL w/v) was added to the reaction mixture from section 2.7.1.1 which contained the reduced TamD (11.23 µg) and incubated for 18 h at 37 ^{O}C. The final enzyme to protein ratio in the digest reaction was ~ 1:20. The reaction sample was then analysed by liquid chromatography-electron spray ionization-mass spectrometry (LC-ESI-MS) as outlined in section 2.10.2.

2.7.1.3 Phosphopantetheinylation Reaction

Recombinant TamD was reacted with malonyl-CoA or CoA as a substrate in an attempt to convert the *apo*-ACP domain into its *holo*-form using AcpS or Sfp (PPTase). TamD (60 μM) was added to reaction buffer J (50 mM ammonium acetate, pH 7.5), with the specific PPTase (1 μM), 1 mM $MgCl_2$, 250 μM CoA or Malonyl-CoA in a final volume of 250 μL. This reaction mixture was incubated overnight at 37 °C followed by LC-ESI-MS analysis.[57]

2.7.2 Reactions for Chapter 4 (SrtA Analysis)

All reactions for this section are specific for SmSrtA unless otherwise stated.

2.7.2.1 Reaction with *trans*-chalcone

SmSrtA (5 μM) was incubated with *trans*-chalcone (0-100 μM) at 4 ^{O}C and a sample removed after 1, 2, 4, 8 and 16 h. The sample incubated with 100 μM of the inhibitor molecule for 16 h was further incubated with DTT (10 mM) for 6 h, followed by dialysis into buffer E for 2 h. All samples were analysed by electrospray ionisation Fourier transform ion cyclotron resonance mass spectrometry (ESI FT-ICR MS) as outlined in section 2.10.1.

2.8 Analysis of Garlic Extracts for Chapter 5

All analyses for this section are specific for BcBCP unless otherwise stated.

2.8.1 Sample Preparation and HPLC Analysis

2.8.1.1 Aqueous Garlic Extracts (AGE)

Garlic was purchased from Sainsbury grocery store and stored at 4 ^{o}C. Equal weights of distilled water (30 mL) and garlic (30 g) were homogenised in a Waring blender to

produce an aqueous garlic extracts (AGE). The homogenate was then strained through a muslin cloth to remove large debris and centrifuged for 40 minutes at 4 ^{O}C at 20 000 rpm and then filter sterilised. The supernatant was then passed through a PD10 column and the small molecular weight fraction was collected and stored at -80 ^{o}C.[159]

2.8.1.2 Aqueous Allicin Standards (AAS)

A stock solution of aqueous allicin standard (AAS) was prepared by dissolving 10 mg of allicin (LKT Laboratories, Germany) in 1 mL of sterile distilled water. Standards that ranged from 8-1000 µg/mL were prepared by serial dilution. All standards were stored at – 80 ^{O} C until further use.

2.8.1.3 AllicinMax™

Four capsules (722 mg) of AllicinMax™ (Allicin International Ltd.) were dissolved in HCl (0.1 M, 1 mL HCl) and the resulting mixture was allowed to shake for 2 h, 200 rpm at 37 ^{o}C. After an hour, a half of the mixture was neutralised with NaOH (0.5 mL, 0.1 M) and then centrifuged for 10 mins at ×13000 g. The sample was diluted with water (dilution factor 10) and filtered (0.45 µm). These steps were repeated with the second half of the original sample after incubation of 2 hr. Both the 1 hr and 2 hr samples were analysed by HPLC using the same method outlined in the quantification of allicin.

2.8.1.4 RP-HPLC Analysis

The allicin content in the AGE was analysed using a published method by Lawson and Wang which is summarized as follows.[160] The RP-HPLC method was first standardised using the AAS standards ranging from 8-1000 µg/mL (Column dimensions: 250 × 4.6 mm, 3.6 µ, C18 Aeris Peptide from Phenomenex). A volume of 20 µL was injected onto the column with an autosampler. The mobile phase composing of MeOH and water

(60:40, v/v) was used for elution with an isocratic gradient at a constant flow rate of 1 mL/min. UV detection at 240 nm was used to analyse eluted fractions.

2.8.2 Reactions with *B. cenocepacia* BCP (BcBCP)

2.8.2.1 BCP and AAS

The purified recombinant BcBCP protein (10 μM) was incubated with AAS (0.1 mM and 1 mM) at 37 ^{o}C for 1, 5, 30, 60, 120 and 180 mins. Each reaction was quenched with 0.1 % formic acid and then analysis by ESI FT-ICR MS. Reversibility of the reaction was determined by incubating BCP (10 μM) with AAS (1 mM) at 37 ^{o}C for 1 hour, then TCEP was added to the reaction mixture to a final concentration of 2 mM.

2.8.2.2 BCP and AGE

Freshly-prepared AGE (3.38 mg/mL was passed through a PD-10 (Sephadex G-25M) desalting column, the low molecular weight fraction was collected and incubated with BcBCP (50 μM) for 1, 5, 30, 60, 120 and 180 mins at 37 ^{o}C. To isolate the modified enzyme from the incubation mixture, the mixture was shaken with equilibrated Ni-NTA resin (1 mL) for 50 mins followed by elution of the protein from the resin with 300 mM imidazole. Excess imidazole was removed from protein samples using a PD-10 column. The eluent was concentrated and analysed by ESI FT-ICR MS.

2.9 Kinetic and Inhibition Studies

2.9.1 FRET Analysis for Chapter 4

2.9.1.1 Basic Principles

Fluorescence or Förster resonance energy transfer (FRET) involves energy transfer between a fluorophore (energy donor) and an energy acceptor or quencher. In order for

FRET to occur, the donor and the acceptor have to be within a distance of 10-100 Å from each other and there has to be a good overlap of the emission spectrum of the donor with the absorption spectrum of the acceptor. The Förster radius (R_0) is the donor-acceptor distance at which the energy transfer is 50%. Within this radius, a transfer of energy between both molecules results in the reduction of fluoresence. As the donor and acceptor get further away from each other, *i.e*. exceeding the Förster radius, there is a reduction in FRET efficiency. This results in the generation of a fluorescent signal which can then be monitored spectroscopically. The application of this principle is a very powerful tool for biological analysis and can be used as a 'spectroscopic ruler' to measure the distance between biological molecules of interest when tagged with a FRET pair.[161]

The FRET assay has been routinely used in measuring the activity of sortases by monitoring the rate at which a flourogenic peptide having the LPXTG recognition motif is cleaved between the threonine and glycine residues. The self-quenching peptide substrate used in this study is 4-([4-(dimethylamino) phenyl]azo)-benzoyl(Dabcyl)-Gln-Ala-Leu-Pro-Glu-Thr-Gly-Glu-Glu-[(2-aminoethyl)-amino]naphthalene-1-sulphonyl (Edans) (*Dabcyl*-QALPETGEE-*Edans*) (Figure 24A). There is very good overlap of the emission spectrum of *Edans* (A_{max} = 340 nm; Em: ~490 nm) and the absoprtion spectrum of *Dabcyl* (A_{max}= ~490 nm). Fluorescence of *Edans* fluorophore within the uncleaved peptide is quenched by *Dabcyl* due to their close proximity described above. Cleavage of the peptide by sortase, between the threonine and glycine residues results in an increased distance between the two molecules and the generation of a fluorescence signal (Figure 24B). The increased fluorescence can be observed as the reaction progresses in a time dependent manner. Emission readings can be monitored at 485-495 nm.

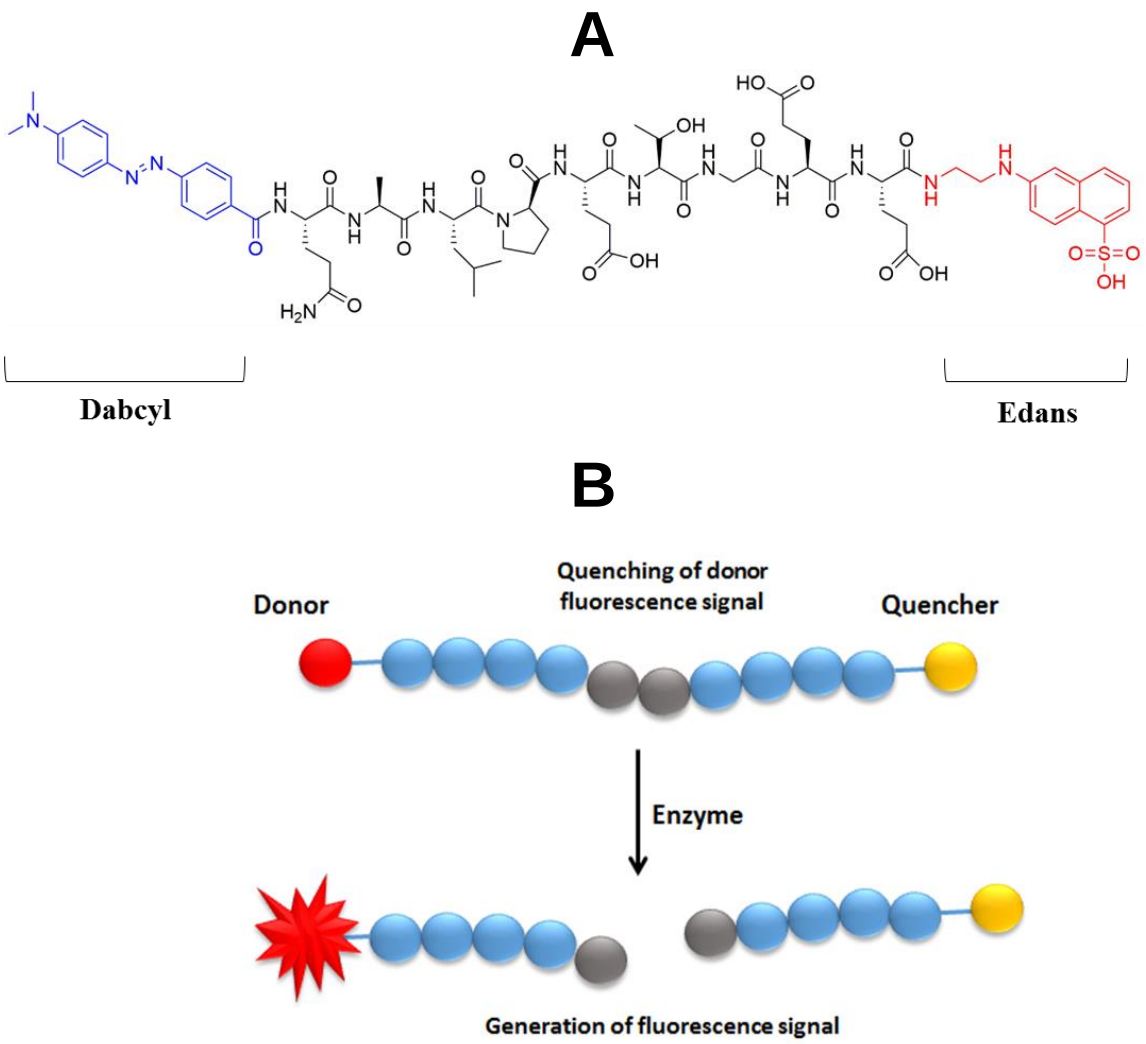


Figure 24: Schematic representation of the proteolytic cleavage a FRET peptide substrate to generate a fluorescence signal. (A) Chemical structure of Dabcyl-QALPETGEE-Edans. (B) Cleavage of Dabcyl-QALPETGEE-Edans. The fluorogenic substrate is self-quenching due to the close proximity of the donor and the acceptor molecules resulting in no fluorescent signal. Upon cleavage at a recongnition site by the enzyme, the distance between the donor and acceptor molecules increases generating a fluorescent signal that can be monitored spectroscopically.

2.9.1.2 Kinetic Analysis of SmSrtA

Dabcyl-QALPETEE-*Edans* was dissolved in dimethyl sulfoxide (DMSO) to prepare a master stock of concentration 3.39 mM. This was then diluted with sterile distilled water so that the final concentration of DMSO in the reaction was kept below 5%. The assay was carried out at 37 ^{o}C in a reaction buffer K (20 mM MES.NaOH, pH 6.5, 125 mM NaCl, 5 mM NH_2-Gly_3 and 2 mM DTT). Assays were performed in a 96-well plate and peptide cleavage was monitored as an increase of fluorescence using a Biotek Synergy HT plate reader. Wells contained 5 µM purified enzyme and varying concentrations of peptide substrate (0.8 -212 µM) in a final reaction volume of 200 µl. The increase in fluorescence intensity was recorded as a function of time using an excitation wavelength at 360 nm and recording emission at 485 nm. Changes in fluorescence were converted to molar velocities using a calibration curve of an equimolar mixture of free *Dabcyl* and *Edans* which had a net gain of 23 F.U./µM upon conversion of substrate to product. The initial velocities for substrate cleavage were determined from the progress curves at various substrate concentrations and were fitted to the Michaelis-Menten equation to determine the kinetic parameters K_m, V_{max} and k_{cat} using GraphPad Prism 6 software. All reactions were performed in triplicates and the mean values reported.

2.9.1.3 Inhibition Analysis of SmSrtA

The inhibitory activity of SmSrtA by *trans*-chalcone was analysed by measuring the rate at which the synthetic fluorescent peptide substrate *Dabcyl*-QALEPETGEE-*Edans* was cleaved in the presence of the inhibitor at various concentrations, using the fluorescent assay described previously in section 2.9.1.2.[155] The reactions were performed on a 200 µL scale in buffer E, triglycine (5 mM), peptide substrate (25 µM), 5 µM recombinant SmSrtA and varying concentrations of *trans*-chalcone (1-100 µM). *trans*-chalcone was dissolved in DMSO and added to reaction mixture so that the final concentartion of DSMO was <1%. Appropriate blanks containing all of the above reagents except for *trans*-chalcone were used as a negative control and morin and curcumin (known natural

product SrtA inhibitors) were used as positive controls. All reactions were done in triplicate.

2.10 Mass Spectrometry (MS)

All MS procedures were performed in collaboration with Dr. David Clarke at the School of Chemistry MS facilities, University of Edinburgh.

2.10.1 FT-ICR MS

General MS method used in chapters 3 to 5.

Protein samples were analysed by LC-MS using an Ultimate 3000 HPLC system (Dionex Corporation, Sunnyvale, CA), equipped with a monolithic PS-DVB (500 µm x 5 mm) analytical column (Dionex Corporation). Mobile phases A and B comprised 2:97.95 and 80:19.95 acetonitrile:water with 0.05% formic acid respectively (v/v/v). Samples were injected onto the analytical column, washed with buffer A for 5 min, followed by a 20 min linear gradient elution (20 µl/min) into buffer B. MS data was acquired on a Bruker 12 Tesla Apex Qe FT-ICR (Bruker Daltonics, Billerica, MA) equipped with electrospray ion source (ESI). Desolvated ions were detected between *m/z* 600 and 2000 for 0.5 s to yield a broadband 512 Kword time-domain data. Fast Fourier Transforms and subsequent analyses were performed using DataAnalysis (Bruker Daltonics) software. All *m/z* spectra were deconvoluted using MaxEnt software in DataAnalysis (Bruker Daltonics).

2.10.2 Peptide Mass Finger-Printing (TamD)

To detect the peptide that contains the proposed lysine residue of TamD that binds PLP, samples of the purified enzyme were prepared and digested as described in section 2.7.1. The resulting peptide mixture was analysed by LC-ESI-MS and the data analysed by DataAnalysis software (Bruker Daltonics). A mass list was created using the SNAP 2.0 algorithm and searched against the known primary sequence of TamD using the MS-Fit

software (University of California). For data searching error tolerances were set to 10 ppm.

2.10.3 Top-Down FT-ICR (SrtA)

Top-down fragmentation was used to determine the amino acid residue of SrtA to which trans-chalcone was covalently bound. Top-down fragmentation was performed on a Bruker 12 Tesla Apex Qe FT-ICR (Bruker Daltonics, Billerica, MA). First the +24 ion species of the chalcone-modified SmSrtA was isolated with the instrument's mass resolving quadrupole. MS/MS was then performed using collision-induced dissociation (CID) or electron capture dissociation (ECD). For CID, the collision voltage was typically set between 20 and 35 V. ECD, 1.8A was applied to the dispenser cathode filament and 20 V to the lens, and a pulse of 4-9 ms was employed. Fragmentation data were the sum of 250-750 scans, and data analyses were performed using DataAnalysis (Bruker Daltonics). The SNAP algorithm was used for automated peak picking, and the resulting top-down fragment mass lists were searched against the primary sequence of SmSrtA using Prosight-PTM software. Mass error tolerances were for all Prosight searches at 10 ppm.

2.10.4 Direct Infusion MS (Allicin)

The mass of the AAS was confirmed using direct infusion mass spectrometry. Typically, AAS was analysed at a concentration of 10 µM from a solution of water/acetonitrile (50:50, v/v). Spectral data were collected on an electrospray micro-ToF mass spectrometer (Bruker) operating in the positive mode.

2.11 Structural Biology

2.11.1 Basic Principles of Protein Crystallisation

X-ray crystallography is a useful tool for the structural determination of proteins, providing a highly detailed picture of the macromolecule at the atomic level. Proteins exist in their native state generally in a fluid environment. In the crystalline state, the molecules self-assemble to form ordered, flexible crystals, held together by weak intermolecular interactions. Crystal stability is a function of building block size and intermolecular properties. Therefore, the inherent flexibility of most proteins and the weak and sparse intermolecular interactions can reduce the likelihood of crystallisation. Additionally, protein molecules have irregular shapes, disordered termini and flexible loops which hinder the protein molecules from stacking and assembling into a regular, periodic lattice. Therefore, many trials are usually required and optimization of crystallization conditions to obtain well-diffracting crystals with good atomic resolution.[162]

The crystallization process begins with a highly concentrated protein solution with precipitants added to reduce protein solubility. The solution then becomes supersaturated and metastable once the solubility limit is exceeded. In the metastable solution collisions between the molecules increase and form transient nuclei, but no spontaneous nucleation occurs (heterogeneous nucleation zone). As super-saturation increases, the nuclei become stable resulting in homogeneous nucleation. These conditions promote the formation of nucleation sites. Crystal growth then continues and stops when equilibrium is reached between the crystals and the saturated protein solution (Figure 25A).[162] Vapor-diffusion is one the most common method to obtain a super-saturated protein solution, using a hanging drop or sitting drop format (Figure 25B). In the vapor-diffusion method, water vapour from a crystallization drop containing protein and precipitant absorbs into a larger reservoir which contains a higher concentration of the precipitant. This process drives the formation of a supersaturated solution which enables crystallization.[162]

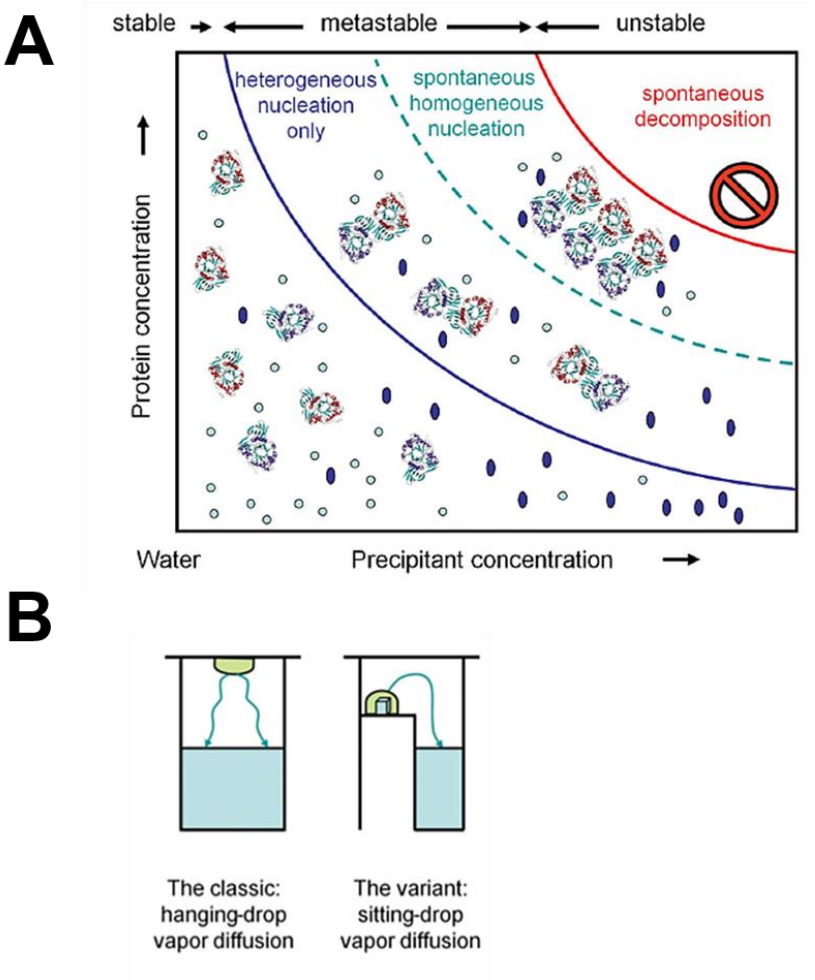


Figure 25: Schematic representation of the protein crystallization. (A) As protein and precipitant concentrations increases a solution that is supersaturated and metastable forms, producing the conditions favourable for homogeneous nucleation/spontaneous crystallisation. (B) Vapour Diffusion method is used to obtain a supersaturated protein solution for crystallisation by either the classic hanging drop or the variant sitting drop set-up. (*Taken from reference 159*)

2.11.2 Structural Studies for Chapter 3 (TamD)

2.11.2.1 Protein crystallization and data collection

Native TamD (after Tev cleavage of hexa-histidine tag) from *P. tunicata* was screened for suitable crystallization conditions at the Scottish Structural Proteomics Facility

(SSPF) in collaboration with Professor J. Naismith and Dr. Stephen McMahon. Protein for crystallisation was concentrated to 24 mg/mL after purification. The crystallization conditions were screened using a nano-drop crystallization robot (Art Griffin). Over 2000 conditions were tested from commercially available sparse-matrix screens using the sitting-drop vapour-diffusion method in a 96-well plate format at 20 ^{O}C. Each well had a volume of 100 µL. Drop sizes of 0.2 µL and 0.3 µL were used in a 1+1 (0.1 µL of protein and 0.1 µL of well solution), 2+1 (0.2 µL of protein and 0.1 µL of well solution) and 1+2 (0.1 µL of protein and 0.2 µL of well solution) format. Several conditions produced crystals and these were optimized. Crystals of TamD were grown over the course of two weeks. The optimum growth conditions are 0.1 M NaOAc, pH 4.6, 40 % (w/v) PEG 2000. Protein to precipitant ratio was 1:1. Prior to data collection the crystals were cryo-cooled in mother liquor doped with 20% glycerol. All crystallographic datasets were collected on beamline I24 at the Diamond synchrotron light source, Oxfordshire, England and processed within the Xia2package in an automated manner using the programs XDS and ScalaData. The output from xia2 was used for solving the structure using PHASER.

2.11.3 Structural Studies for Chapter 4 (SrtA)

2.11.3.1 Protein crystallisation and data collection

The H139A mutant *S. mutans* SrtA (ΔN40 truncated construct) was used for structural studies in collaboration with Dr. Jon Marles-Wright at the Institute of Structural and Moleuclar Biology, University of Edinburgh. SrtA was concentrated to 11 mg/ml after purification. Concentrated protein was crystallised by hanging drop vapour diffusion in drops of 2 µl protein plus 2 µl crystallisation solution (30 % w/v PEG 4000, NaOAc pH 4.6, 0.2 M $(NH_4)_2SO_4$), over 1 ml of the latter. Crystals were obtained in 5 days grown at 20 ^{O}C; these were harvested from the well using a LithoLoop (Molecular Dimensions Limited), transferred briefly to a cryoprotection solution containing well solution supplemented with 20 % w/v PEG200, and the crystals were subsequently flash cooled in liquid nitrogen. All crystallographic datasets were collected on beamlines I24 and I04

at Diamond Light Source (Didcot, UK) at 100 K using Pilatus 6M detectors. Diffraction data were integrated and scaled using XDS and symmetry related reflections were merged with Aimless.[128]. The resolution cut off used for structure determination and refinement was determined based on the CC1/2 criterion proposed by Karplus and Diederichs.[163]

2.11.3.2 Structure solution and analysis

The structure of SmSrtA_H139A mutant was determined by molecular replacement using *S. pyogenes* SrtA (PDBID: 3FN5)[136] as the search model, this was modified to match the sequence of the target protein using Chainsaw.[164] A single solution comprising a monomer in the asymmetric unit was found using Phaser.[165] The initial model was rebuilt using Phenix.autobuild[166] followed by cycles of refinement with Phenix.refine[166] and manual rebuilding in coot[167]. The final model was refined with anisotropic B-factors for the protein chain and isotropic B-factors for ligands and water molecules. The model was validated using MolProbity[168]. Structural superimpositions were calculated using Coot. Crystallographic figures were generated with PyMOL.[169]

2.11.3.3 Molecular Modelling

The molecular modelling work was carried out in collaboration with Dr. Jon Marles-Wright. Both the R- and S- form of the proposed *trans*-chalcone cysteine Michael adduct on Cys205 of SrtA were built and their geometry was optimized using Jligand.[170] Geometry description files were output as cif format files and Cys205 was modified in coot with both forms of the trans-chalcone. The His139 residue was modeled to minimize steric clashes with other side-chains and the ligand. The resulting models were minimized using refmac and figures produced using PyMol.[169]

2.12 Antimicrobial Assays

2.12.1 Antimicrobial Analysis for Chapter 4

2.12.1.1 The ACTA Active Attachment Biofilm Model

The ACTA active attachment (AAA) model is a novel high-throughput screening model for testing preventive compounds against oral pathogens. The model was developed by the Academic Centre for Dentistry in Amsterdam in 2010. The assembly consists of a stainless steel lid with 24 plastic clamps that contain glass cover slips (Figure 26). This is used as a substratum for growing biofilms after being conditioned with sterile saliva. The assembled lid is able to fit onto a standard Corning flat-bottom 24-well plate.[170]

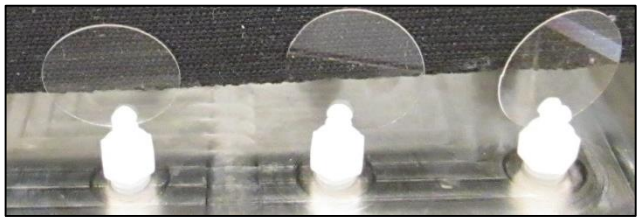


Figure 26: Assembled lid with glass slides used in AAA model for biofilm growth. Glass slides are sterilized and used as a substratum for growing biofilms after being conditioned with sterile saliva.

2.12.1.2 Biofilm Growth

The antimicrobial assays for *S. mutans* were performed in collaboration with Dr. Amarnath Maitra at Wrigley's Global Innovative Centre, Chicago. *S. mutans* culture was grown overnight in 10 mL of brain heart infusin broth (BHI) with or without actives/inhibitors at 37 °C. The culture was diluted to a final optical density (OD_{600}) of 0.05 for use as an inoculum. The assembled AAA apparatus was incubated with 1.5 mL of sterile 50-fold diluted saliva for 2 h at 37 °C. The glass discs were inoculated with *S. mutans* culture (OD_{600} = 0.05) grown in 1.5 mL of BHI containing 0.2% sucrose for 2.5 h with moderate shaking (100 rpm). After initial inoculation, the assembled lid was

placed in fresh media (without bacteria) and biofilm was grown for a period of 18 h. The *S. mutans* culture was treated with the actives/inhibitors at every stage of growth (planktonic to biofilm) at concentrations between 0-500 µM.

2.11.1.3 Biofilm Growth Analysis

After 18 h of biofilm growth, the glass discs with the biofilm were transferred into a 24-well plate containing 1.5 mL of 0.05% crystal violet solution for staining up to 10 mins. The discs were then washed with 2 mL of phosphate-buffered saline (PBS) and allowed to air dry. They were then de-stained in 2 mL of 95% ethanol and the absorbance of each well solution (corresponding to a glass disc) was then measured at 600 nm. The absorbance of the well solution was taken to be directly proportional to biomass/biofilm formation.

2.12.2 Antimicrobial Assays for Chapter 5

The assays performed in this section with AGE and AAS were carried out in collaboration with Professor John Govan and Dr. Cathy Doherty at the Centre of Infectious Diseases, School of Medicine, University of Edinburgh.

2.12.2.1 Agar Dilution

The MICs of AGE against the 38 isolates were investigated by the agar dilution method as recommended by the National Committee for Clinical Laboratory Standards (NCCLS). In summary the bacterial strains were cultured overnight at 37 ^{O}C in Muller-Hinton broth (MHB) to obtain a bacterial inoculum of 10^6 CFU/mL. Muller-Hinton agar (MHA) plates were prepared by adding AGE at concentrations between 0.1-10 % AGE from concentrated stock solution (AGE neat stock was made from 1g crushed garlic per 1 mL of distilled water). Agar plates were incubated overnight at 37 ^{O}C and examined for growth the following day. The MIC is taken as the lowest concentration without visible growth.

2.12.2.2 Microtitre Broth Dilution

The MICs and MBCs for AAS against representative members of the strain panel were determined by microtitre broth technique recommended by the NCCLS.[171] In summary MHB (50 µL) was aliquoted to each well of a 96-well microtitre plate. AAS (50 µL) was added to the first well and a 2-fold serial dilution performed in subsequent wells (concentration range: 0.125-128 mg/mL for AAS). Standardized broth inoculum (10^6 CFU/mL) was added to each well. An absorbance reading at 625 nm was recorded for each well pre- and post-incubation at 37 ^{O}C for 16-18 h. The lowest concentrations of AAS which showed no increment of absorbance readings after incubation were considered as the MIC values. The lowest concentrations which did show any bacterial growth upon spotting on MHA plates after overnight incubation at 37 ^{O}C were considered as the MBC values.

2.12.2.3 Agar Well and Paper Disc Diffusion

The susceptibility of *B. cenocepacia* C3168 to AGE was demonstrated by agar well diffusion and impregnated paper disc diffusion analysis.[159] Isosensitest agar (Oxoid Ltd., Basingstoke, UK) plates were flood seeded using a 10^6 CFU/mL isosensitest broth culture of C3168. 10 µL undiluted AGE was added to a well cut in the agar or used to impregnate sterile paper discs. Plates were then incubated at 37 ^{O}C for 16-18 h followed by the examination of zones of inhibition by eye.

Chapter 3: Structural Analysis of a TamD

3.1 Introduction

3.1.1 TamD

The *tam* cluster identified in the genome of the marine bacterium *P. tunicata* contains 21 ORFs that encode 21 protein products, 19 of which are involved in the biosynthesis of the tambjamine, YP1. Kjellberg and colleagues cloned this cluster in *E. coli* and showed that it was capable of producing YP1 in this recombinant host.[39] Among the genes in this cluster is the *tamD* gene which encodes a putative AOS enzyme, TamD. The TamD protein comprises of 536 amino acids and sequence analysis and homology predict that the enzyme exists as a natural fusion protein, with an ACP domain at its N-terminus and a PLP-binding AOS domain (Figure 27). This arrangement is unusual for the AOSs, as their associated ACP partners are usually expressed as discrete proteins, e.g. the *S. wittichii* SPT and its ACP partner.[57] An analysis of the tambjamine biosynthetic pathway suggests that TamD is involved in the incorporation of a malonyl unit via the ACP domain and a seryl unit from L-serine via the AOS domain to produce a bipyrrollic intermediate by a catalytic PLP-dependent condensation reaction (Figure 3).

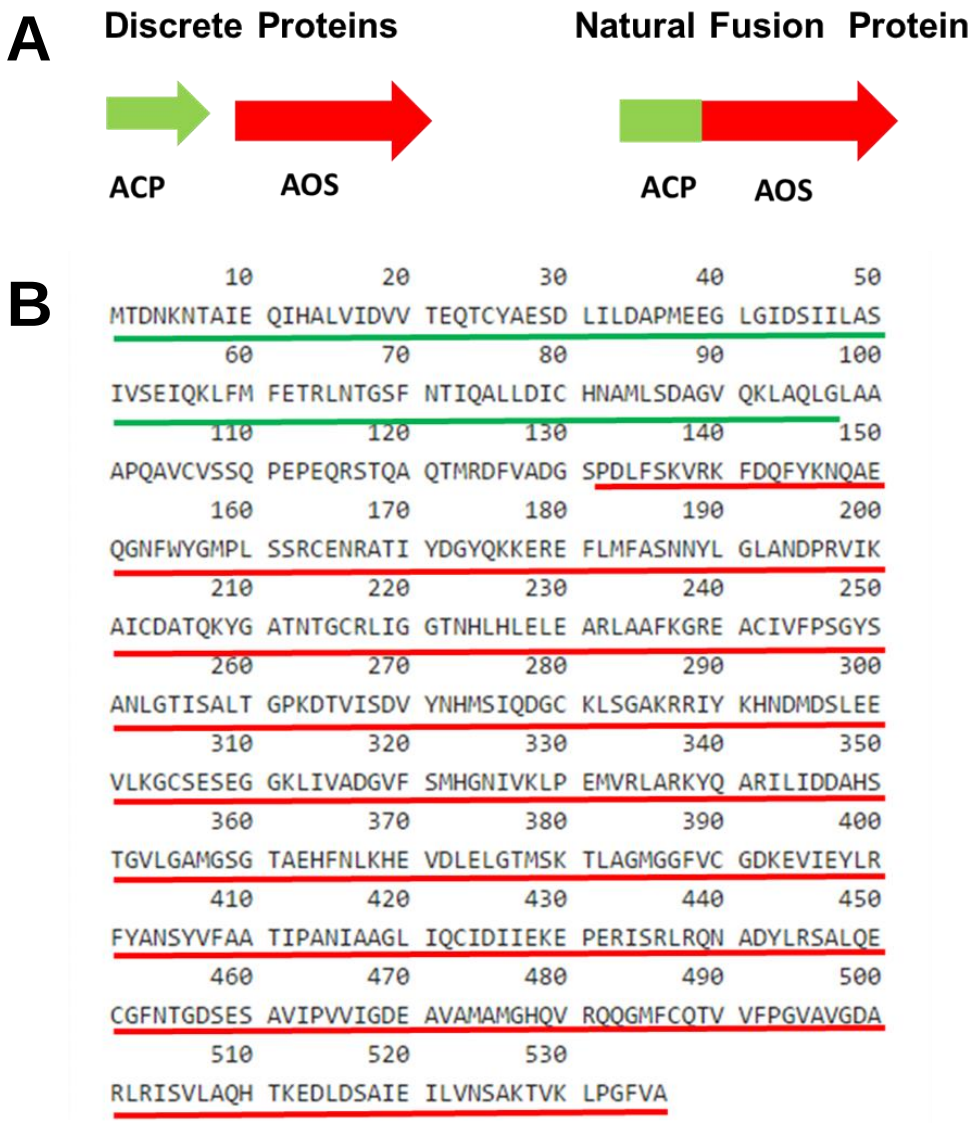


Figure 27: Schematic representation of TamD primary sequence. (A) The typical gene organisation of AOS and ACP resulting in discrete protein products.[57] The *tamD* gene from tambjamine biosynthetic pathway which encodes for the TamD protein produces a natural fusion protein with an ACP and an AOS domain. (B) The primary sequence of TamD has a predicted ACP domain at the N-terminus (underlined in green) separated from the predicted AOS domain (underlined in red) by a short linker region.

The AOS domain of TamD is expected to have the conserved structure observed in the other members of the AOS family determined to date. These include SPT[56, 57], AONS[50], ALAS[59], KBL[55] and more recently CqsA[53]. These enzymes exist functionally as

symmetric homodimers belonging to the α-group and the fold type 1 superfamily of PLP-dependent enzymes (Figure 28).[65] They also share a similar topology, with each monomer consisting of three domains: an N-terminal domain, a central catalytic domain which binds PLP and a C-terminal domain. The N-terminal domain consists of an α-helix and a three-stranded antiparallel β-sheet, while the central domain is made up of a seven-strand β-sheet, usually flanked on both sides by α-helices. The residues of the C-terminal domain interact closely with both the N-terminal and the catalytic domains of the same monomer. Both monomers are tightly interlocked with residues from each monomer involved in dimer formation.[50, 53, 55, 56, 59]

The sequence alignment of TamD with AOSs shows high conservation of the key catalytic residues in the active site and those responsible for PLP binding, which suggest a common reaction mechanism and active site geometry (Figure A3). Each monomer binds a PLP molecule forming a Schiff base/internal aldimine via a conserved lysine residue in the active site. Other residues which are involved in PLP binding include His216 and Asp213 which stabilize the pyridinium ring through hydrogen bonding and salt bridge formation respectively. Additionally His139, which is found at the beginning of a three-residue motif (His139-Ala140-Ser141) stacks on one face of the pyridinium ring through hydrophobic interactions in all AOS enzymes.

The ACP domain of TamD has very low sequence homology to both FAS and PKS ACPs (Figure A4). A BLAST search from the PDB suggests that TamD-ACP has the closest structural homology to *B.subtilis* ACP (PDB code 1HY8), although they have a 21% similarity in sequence indentity.[172, 173] This unique feature of the TamD enzyme has sparked interest into further investigations of the topological arrangement and protein-protein interactions of the ACP and the AOS domains. Information from the three-dimensional structure of TamD will be useful in providing insights into the process and mechanism involved in the sequestering of acyl-ACP substrates from the ACP to the active site of AOS enzymes.

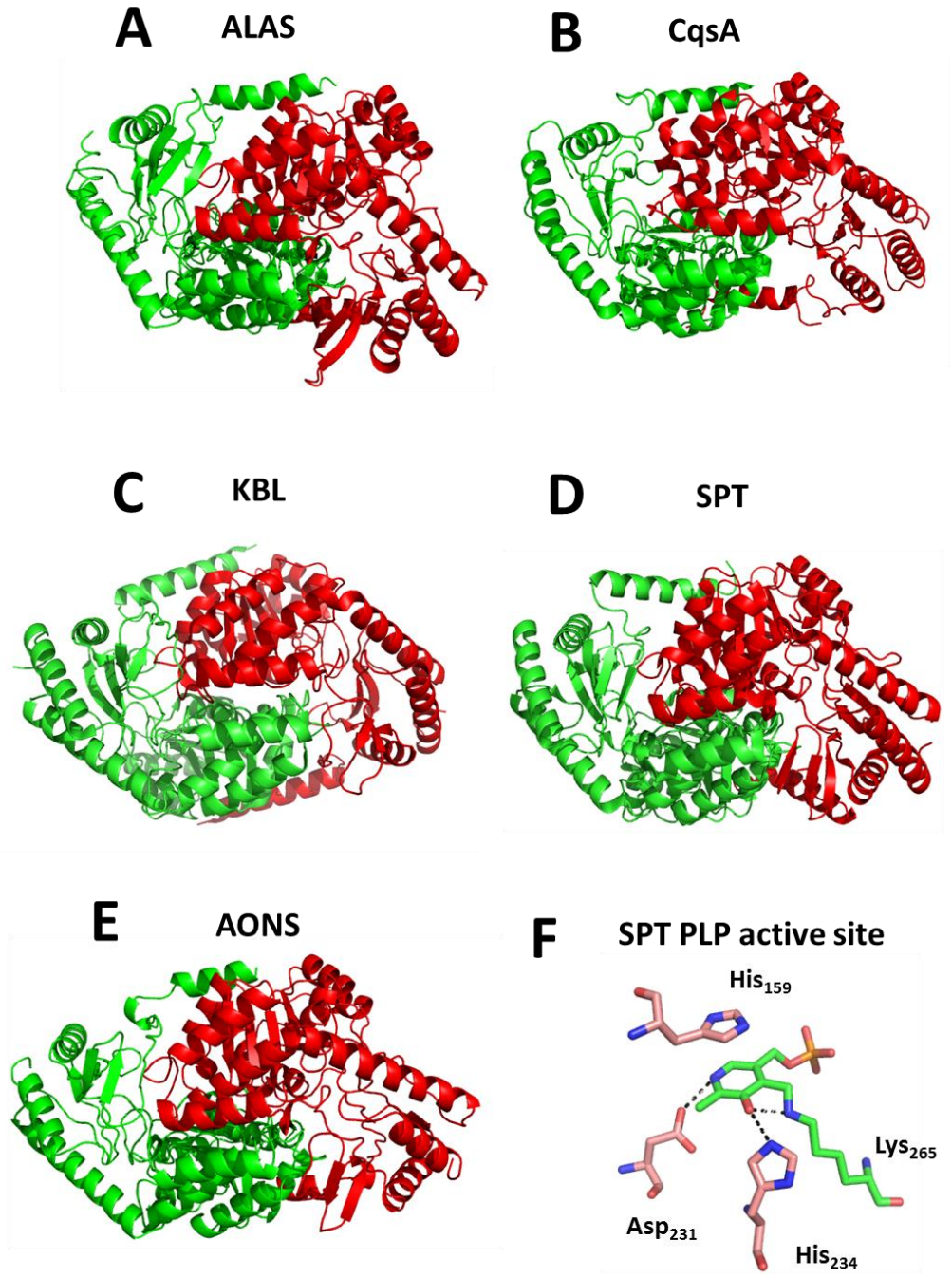


Figure 28: Crystal Structures of AOSs. (A) *Rhodobacter capsulatus* ALAS (PDB code: 2BWN); (B) *V. cholerae* CqsA (PDB code: 2WK8); (C) *E. coli* KBL (PDB code: 1FC4); (D) *Sphingomonas paucimobilis* SPT (PDB code: 2JG2); (E) *E. coli* AONS (PDB code: 1BS0) and (F) Key residues involved in PLP binding of AOS enzymes.

3.1.2 Aims

The structure of the natural fusion protein TamD is very crucial to the understanding of the biophysical interactions between ACPs and their AOS partners. The aim of this study was to characterize TamD and determine the three-dimensional structure of its *apo*, *holo* and acylated forms by x-ray crystallography. With the dynamic nature proposed for TamD and the predicted high disorder in the N-terminal region where the ACP domain is located, another aim was to optimise suitable crystallographic conditions to obtain a high resolution x-ray crystal structure of the enzyme.

3.2 Results and Discussion

3.2.1 Cloning *tamD* from Genomic DNA

The *tamD* gene was amplified from the genomic DNA of *P. tunicata* using forward and reverse primers appended with *Nde1* and *Xho1* restriction sites respectively to give a PCR product of ~1.6 kpb (Figure 29A). The gene was sequenced for confirmation and cloned into the pET22b expression vector that would translate into the TamD protein with a C-terminal hexa-histidine tag. Additionally, the gene was also cloned into a pHISTEV plasmid via the *Nco1* and *Xho1* restriciton sites to obtain a translated protein product with a cleavable N-terminal hexa-histidine tag. The pHISTEV plasmid has an N-terminal TEV protease site located downstream from the hexa-histidine site (Figure 23), therefore upon cleavage with TEV protease, a native protein would be obtained.

Since the *tamD* gene encodes for a didomain protein (ACP and AOS domains), attempts were made to amplify both domains separately. The ACP domain is estimated to be approximately 280 bps in length from the start of the gene, while the remaining gene product is expected to encode the AOS domain and a short disordered linker region, altogether being approximately 1.5 kbp in length based on sequence homology with the

other ACPs (Figure 29B). These gene fragments were subsequently cloned into pET11a and pET22b expression vectors, respectively, after amplification.

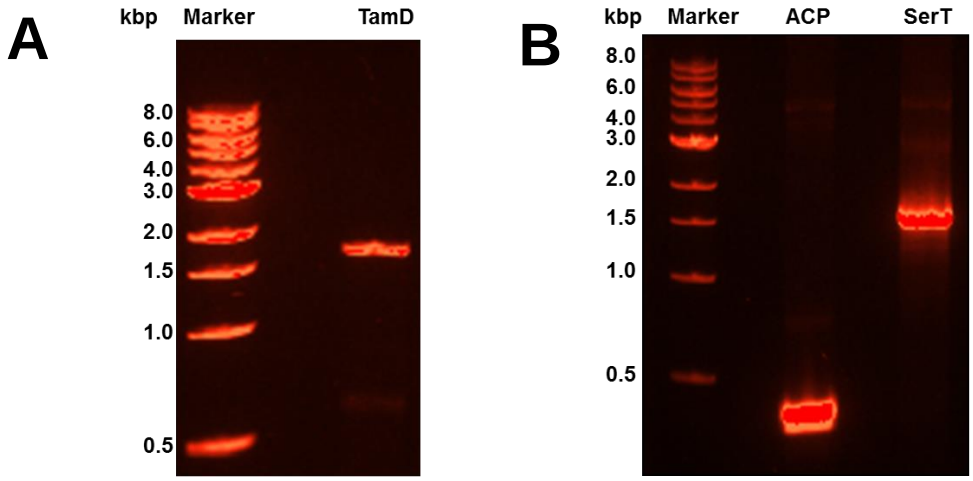


Figure 29: Amplification of *tamD* gene from *P. tunicata* genomic DNA (A) PCR product of *tamD* gene which is ~1.6 kbp in length. (B) PCR products from the amplification of the gene squences encoding ACP and the AOS domains of TamD which are ~ 280 bp and ~1400 bp in length.

3.2.2 Protein Expression and Purification

3.2.2.1 TamD/pET22b Construct

Full-length *tamD* was expressed in an *E. coli* host, strain BL21 (DE3) to give a C-terminal hexa-histidine tagged protein, after induction with 0.1 mM IPTG and incubating for 4.5 hours at 20 ^{o}C. The protein was isolated from cell free extract (CFE) by nickel affinity chromatography eluting with increasing imidazole concentration. SDS/PAGE analysis of the elution fractions under reducing conditions resulted in a band at ~59 kDa which corresponding to the TamD monomer (Figure 30A). Further purification and oligomeric analysis by size exclusion chromatography (SEC) gave a symmetric peak corresponding to the theoretical mass of ~ 120 kDa which suggests that TamD exists in solution as a dimeric protein (Figure 30B).

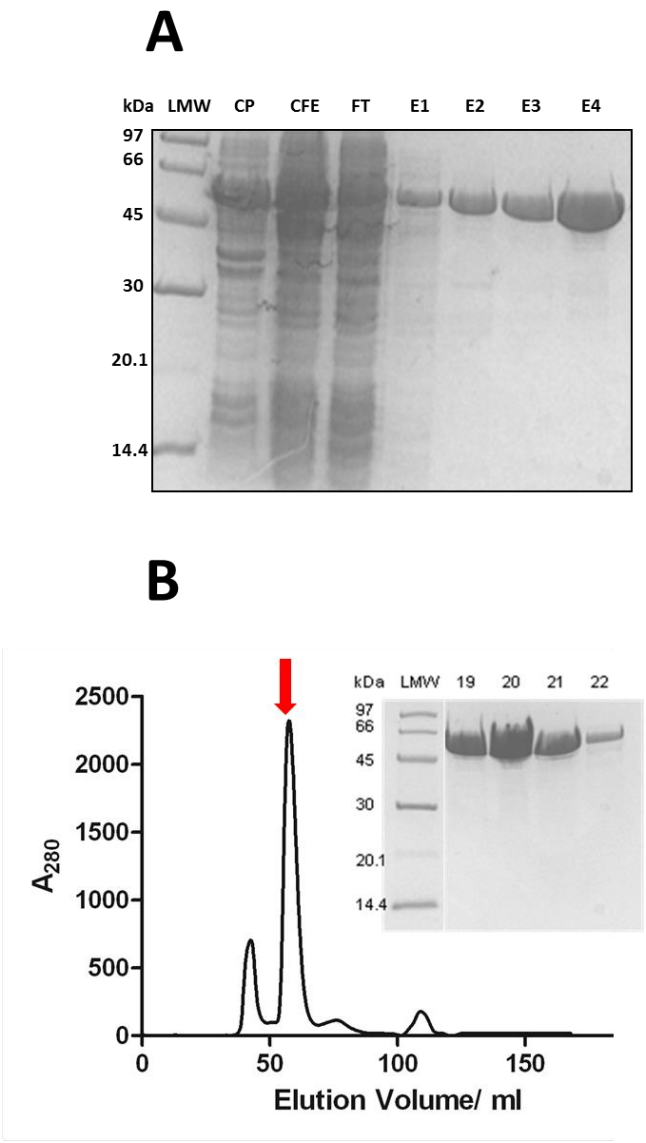


Figure 30: Purification of Full-Length TamD with C-terminal His-tag. (A) SDS-PAGE analysis TamD purified by Ni-NTA Chromatography. LMW: low molecular weight marker; CP: cell pellet; CFE: cell free extract; FT: flow through E1: elution with 20 mM imidazole; E2: elution with 40 mM imidazole; E3: 60 mM imidazole; E4: 200 mM imidazole. TamD present in elution fractions. (B) Elution profile of TamD purified by SEC on Superdex S200 16/60.TamD is highlighted with red arrow. Elution fractions analysed by SDS/PAGE to give a band at ~59 kDa corresponding to the monomer (*inset*).

3.2.2.2 TamD/pHISTEV Construct

The TamD with an N-terminal HISTEV tag was expressed as previously described in section 3.2.2.1. The advantage of using this construct for expression is that it allows for removal of the His-tag after purification by cleavage at the TEV protease site resulting in the production of native protein. The purification process involved an initial isolation of the HISTEV-tagged TamD from the CFE by nickel affinity chromatography using the same protocol described in the section above. The tag was cleaved by overnight incubation of the protein at 4 ^{O}C with TEV protease using enzyme to protease ratio of 10:1. The N-terminal His-tag and TEV protease were subsequently removed from the protein sample, by a second nickel affinity purification step (Figure 31A), giving a final yield of ~70 mg/mL per 5L of culture. Further purification by SEC resulted in the protein eluting in a similar manner to the C-terminal tagged version (Figure 31B).

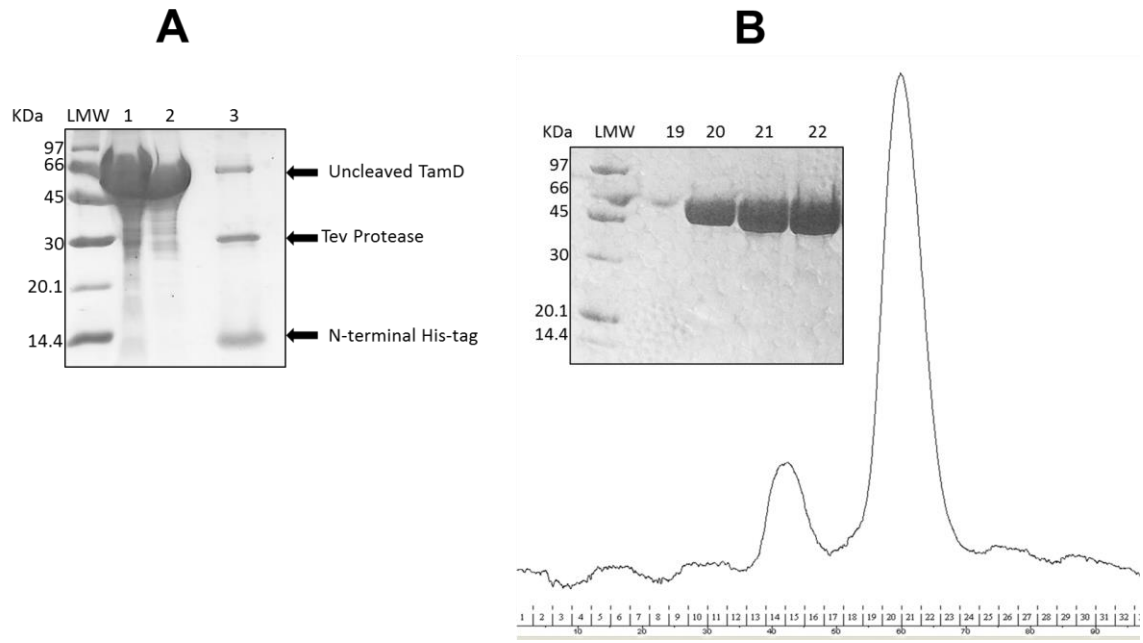


Figure 31: Purification of TamD with N-terminal His-tag. (A) SDS-PAGE analysis TamD purified by Ni-NTA Chromatography. LMW: low molecular weight marker; 1: TamD before Tev protease cleavage; 2: TamD after Tev protease cleavage; 3: His-tagged fraction eluted from resin after cleavage reaction. (B) Elution profile of TamD purified by size exclusion chromatograpy on Superdex S200. Elution fractions analysed by SDS/PAGE to give a band at ~59 kDa corresponding to the monomer (*inset*).

3.2.2.3 TamD-ACP and TamD-AOS Constructs

The TamD-AOS/pET22b construct was expressed in *E. coli* (BL21 DE3) and the AOS domain was purified in the same manner as the intact protein described in section 3.2.2.1 by nickel affinity chromatography and SEC. The protein eluted as a dimer with a monomeric mass of ~ 49.5 kDa as confirmed by SDS-PAGE analysis and LC-ESI-MS yielding ~42 mg/mL per 5L of culture. The expression and purification of the TamD-ACP construct was unsuccessful. This could be due to the fact that the protein was unstable without its AOS domain present. Therefore new constructs with less/more truncation will have to be explored in order to find a suitable fragment which can express the ACP domain as a stable soluble protein.

3.3.3 Characterisation by UV/vis Spectroscopy

3.3.3.1 Spectroscopic measurement of PLP-bound TamD

Purified TamD has a yellow colour due to the presence of bound PLP, which can be analysed spectroscopically as PLP absorbs within the UV/visible range (Figure 32A). After converting all of the isolated protein to the *holo*-form by incubation with PLP, excess cofactor was removed by passing the enzyme through a desalting PD-10 column before spectroscopic measurements. The absorption spectrum was examined by scanning the protein sample from 200-800 nm. The TamD spectrum revealed prominent absorption maxima at 330 nm and 426 nm corresponding to the absorbance of the internal aldimine which is formed from PLP covalently bonding to the Lys380 residue of TamD (Figure 32B). This observation correlates well to other AOS PLP-dependent enzymes which have characteristic spectral peaks of 330-340 nm and 415-426 nm which corresponds to the tautomeric forms of the bound PLP cofactor.[174] The enolamine form (λ_{max} 330 nm) and the ketoenamine (λ_{max} 426 nm) existing in equilibrium, with the latter form being more dominant in TamD. Upon addition of L-serine, the amino acid substrate of TamD, there is the formation of an external aldimine with PLP which results in an increase in absorbance at ~426 nm and a spectral shift to left of ~2 nm (Figure 32B). This results in a displacement of the active site lysine residue.

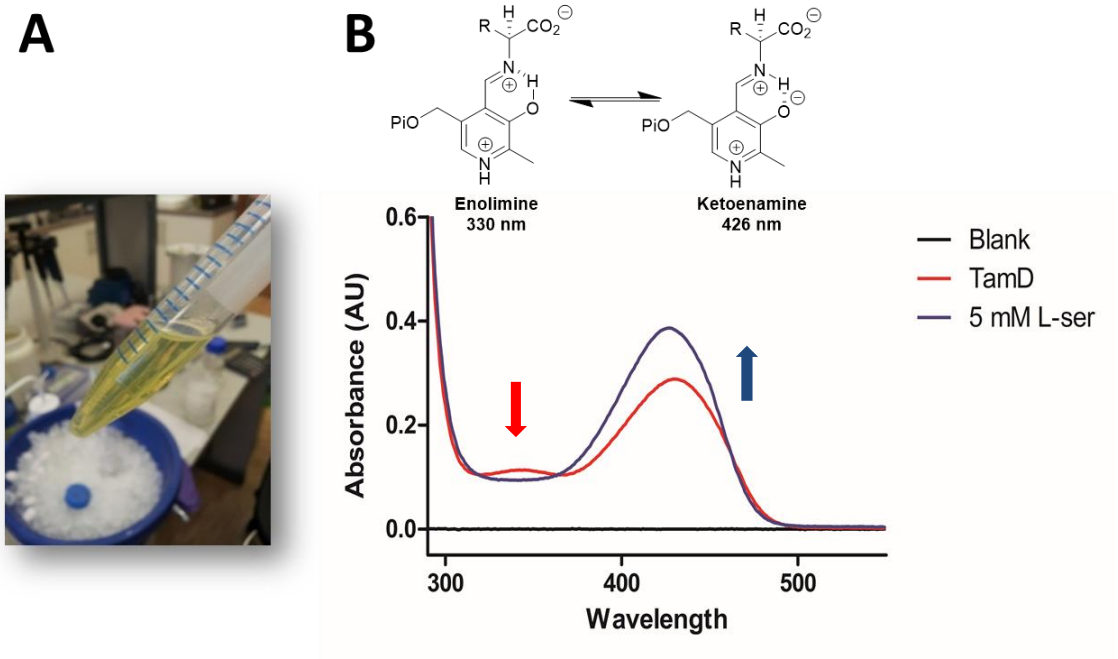


Figure 32: PLP-bound TamD after purification. (A) Photograph showing that TamD has a characteristic yellow colur after purification and removal of excess PLP by PD-10 desalting column. This is due to the presence of the bound PLP cofactor via a Schiff base to the active site lysine residue. (B) The UV/vis spectrum shows absorbance maxima at 330 nm and 426 nm corresponding to the enolamine and ketoenamine tautomeric forms of PLP respectively, with the latter being more dominant in TamD. Spectra were recorded on Varian's Cary 50 UV/vis spectrophotometer and the Cary WinUV software used for analysis of results. Samples were analysed in a quartz cuvette with a path length of 1 cm from 200-700 nm.

3.3.3.2 Binding Constant of L-serine

The ability of TamD to bind L-serine and form an external aldimine was measured spectroscopically by adding increasing concentrations of L-serine (ranging between 0 and 100 mM) to the enzyme (25 µM). This addition resulted in an equilibrium shift to the ketoenamine form of PLP with a distinct increase in the peak at 426 nm and a shift to the left by ~2 nm. By measuring the changes in absorbance at the ketoenamine-specific wavelength, the apparent L-serine binding constant (K_d) was determined to be 5.01 ± 0.64 mM (Figure 33). This value is comparably higher than the K_d of L-serine for SPT

(K_d=1.1 mM), another well characterized AOS enzyme which utilizes L-serine as a substrate.

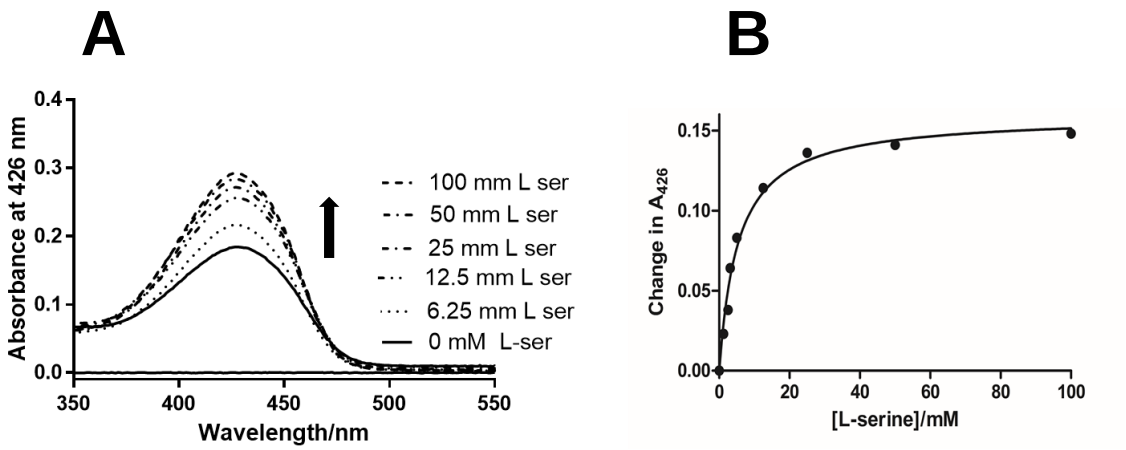


Figure 33: Binding of L-serine to TamD. (A) UV/vis spectroscopic measurements confirmed that L-serine is a substrate of TamD forming an external aldimine with the enzyme resulting in spectral changes that can be monitored at 426 nm. (B) Plot showing dissociation constant of L-serine for TamD. Scans were done with a Varian Cary 50 UV/vis spectrophotometer and the Cary WinUV software used for analysis of results. Samples were analysed with quartz cuvettes with a pathlength of 1 cm from 200-700 nm.

3.3.4 Characterisation by Mass Spectrometry

3.3.4.1 TamD Analysis by ESI-MS

Purified TamD was analysed by LC-ESI-MS under reducing conditions. The mass spectrum confirmed the mass of TamD to be 59516 Da, this mass was +2 Da higher than the predicted mass of recombinant TamD based on the amino acid sequence (59514 Da). Another peak at 59385 Da was also observed which corresponds to the loss of the N-terminal methionine (Figure 34). It was not possible to obtain a more accurate mass of the protein due to its large size. The observed mass was representative of the *apo*-form of the enzyme, with no bound PLP detected. Therefore further MS analysis was carried out to determine the type and site of PTM between PLP and TamD.

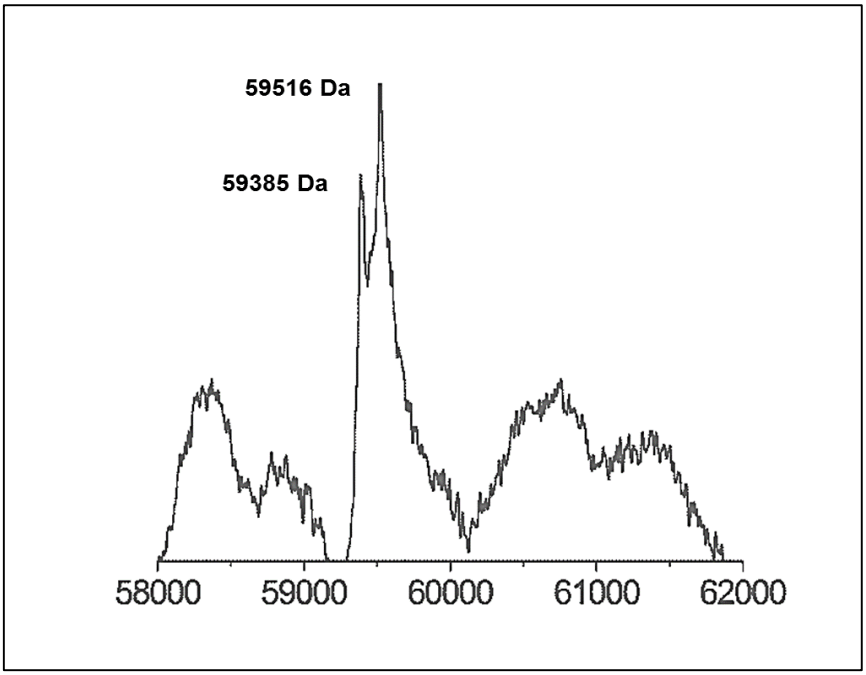


Figure 34: Positive ion mass spectrum of purified tamD. Two peaks were observed for *apo*-TamD expressed with and without the N-terminal methionine which was cleaved after expression giving masses of 59516 Da and 59385 Da respectively.

3.3.4.2 Peptide Mass Fingerprinting

The primary sequence of TamD was confirmed by MS analysis after fragmentation in solution with trypsin and LysC. The observed peptides covered ~80% of the protein sequence (Figure A5). In order to identify the lysine residue of TamD which forms an internal aldimine with PLP, $NaBH_4$ was used to reduce the imine bond formed between the amino group of the lysine side chain and the aldehyde of PLP to an amine (Figure 35A). This reaction was monitored spectroscopically with a characteristic reduction of the absorbance maxima at 426 nm and the formation of a new peak at 330 nm. The reduction reaction was followed by digestion of the reduced PLP:enzyme adduct in solution with LysC endoproteinase, which cleaves the protein after every lysine residue (Figure 35B). Analysis of the peptide fragments by MS identified the peptide fragment cleaved between Lys369 and Thr381 (HEVDLELGTMSK) to have a mass increase of

+231.03 Da which corresponds to the mass of the reduced PLP adduct (Figure 36). This confirmed Lys380 as the active site residue which forms an internal aldimine with PLP.

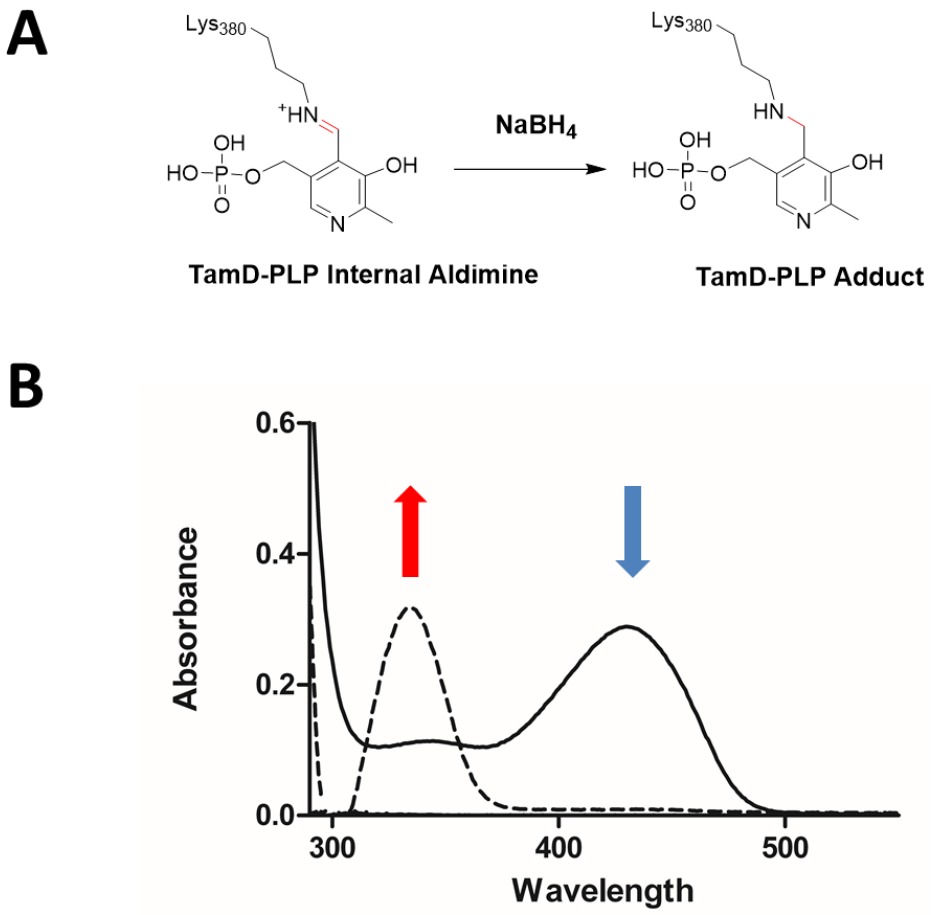


Figure 35: LysC digestion and analysis of TamD after reduction with sodium borohydride. (A) Reaction scheme for the reduction of TamD-PLP aldimine with $NaBH_4$ to form an amine adduct. (B) UV/vis spectrum of TamD-PLP aldimine reduction. TamD has absorbance maxima at 426 nm and 334 nm due to the PLP cofactor which forms an internal aldimine with the active site lysine (solid black line). The addition of $NaBH_4$ (1 mM) results in a covalent TamD-PLP adduct with a characteristic increase of the peak at 330 nm (dotted trace) and reduction of the ketoenamine peak at 426 nm after 15 mins.

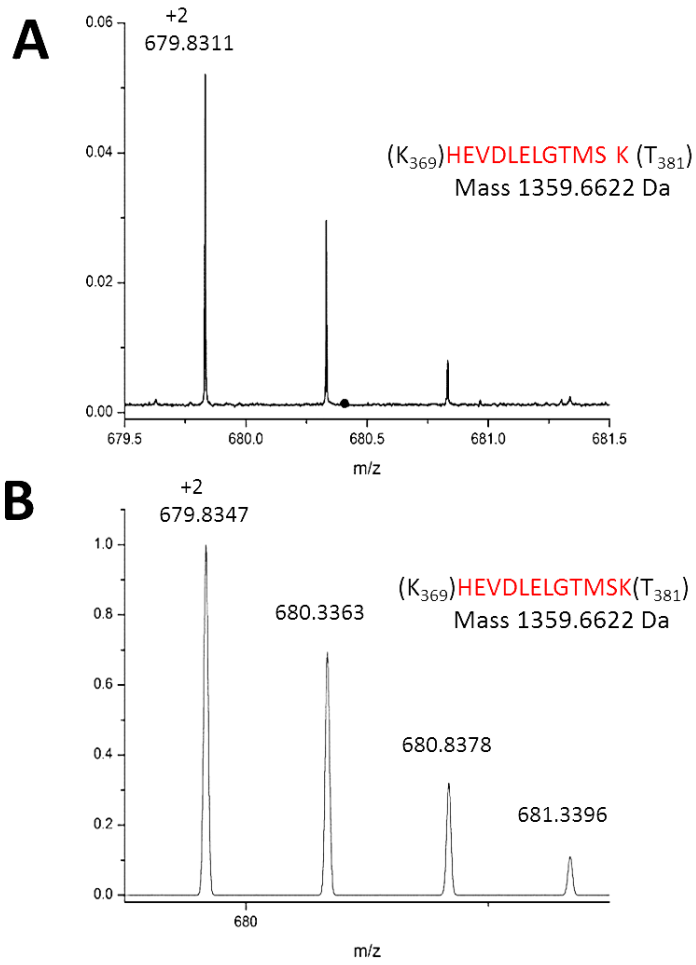
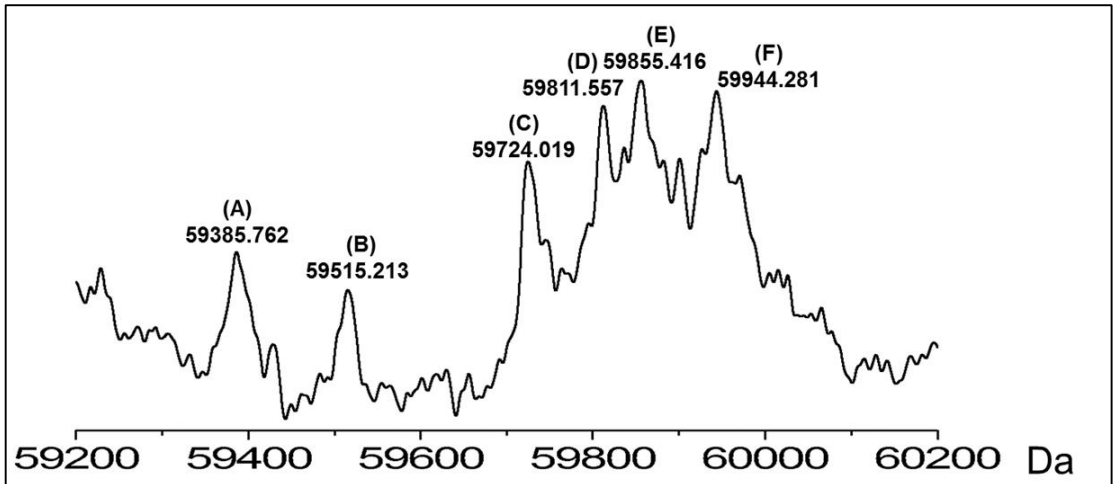


Figure 36: Mass spectrum of TamD peptide fragment with PLP adduct. (A) Mass spectrum of peptide consisting of amino acids 370 to 381 with PLP adduct: +2 charge state, m/z = 679.8311 and average mass 1359.6622 Da. (B) Simulated mass spectrum of theoretical mass of peptide fragment of residues 370 to 381 with PLP adduct. Isotopic envelop of the theoretical mass corresponds to that of the observed mass.

3.3.4.3 Analysis of phosphopantetheinylation of TamD

The MS analysis of TamD suggested that the enzyme is expressed with its ACP domain in the *apo* form. Attempts were therefore made to convert the TamD-ACP domain to its *holo* form using a modified version of the method described previously by Raman which was used to convert *S.wittichi apo*-ACP to its *holo* form.[57] In summary, TamD was

incubated with *Bacillus* Sfp, $MgCl_2$, DTT and malonyl-CoA/CoA overnight at 37 ^{O}C. The acylation reaction was then analysed by LC-ESI-MS. The resulting spectrum from the reaction mixture was very complex indicating several species present in solution which included the *apo*, *holo* and the acylated forms of TamD-ACP. All three forms of TamD were identified with and without the N-terminal methionine (Figure 37). The use of *E. coli* AcpS in the acylation reaction did not result in any modification of the *apo* TamD-ACP, which suggests that the NRPS PPTase (Sfp) was more readily recognized as a partner for TamD-ACP than the FAS PPTase (AcpS). In addition to the attempts to acylate TamD *in vitro*, the enzyme was also coexpressed *in vivo* with *Bacillus* Sfp with the aim of obtaining acylated TamD upon purification. MS analysis of purified TamD after co-expression with Sfp resulted in *holo* TamD with and without the N-terminal methionine giving masses of 59855.460 Da and 59724.011 Da respectively (Figure 38).



Peak	Observed Mass (Da)	Species
A	59385.762	Apo TamD-ACP (minus N-terminal Met)
B	59515.213	Apo TamD-ACP
C	59724.019	Holo TamD-ACP (minus N-terminal Met)
D	59811.557	Malonylated TamD-ACP (minus N-terminal Met)
E	59855.416	Holo TamD-ACP
F	59944.281	Malonylated TamD-ACP

Figure 37: Positive ion mass spectrometry of TamD after acylation reaction. *Apo* TamD was phosphopantetheinylated by *Bacillus* Sfp in the presence of Mg^{2+} and malonyl-CoA. The complex mass spectrum shows species expressed with and without the N-terminal methionine in the *apo, holo* and acylated forms.

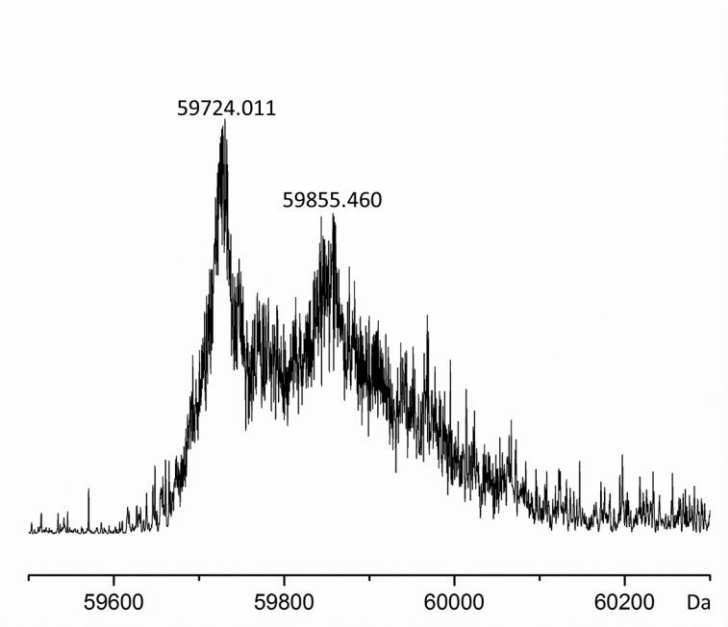


Figure 38: Positive ion mass spectrometry analysis of purified TamD co-expressed with Sfp. TamD was coexpressed *Bacillus* Sfp to produce *holo* enzyme with and without the N-terminal methionine giving masses of 59855.460 Da and 59724.011 Da respectively.

3.3.5 Structural Determination of TamD

3.3.5.1 Crystallisation

The structural study of TamD was done in collaboration with Professor J. Naismith and Dr. Stephen McMahon in St. Andrews. TamD was concentrated to 24 mg/mL prior to crystallisation trials. The crystals were grown by vapour drop diffusion at 20 °C over the course of two to three weeks (Figure 39). The optimal conditions for crystal growth were 18.6 % PEG 3350 and 0.7 M $MgSO_4$ at a protein to precipitant ratio of 2:1. The crystals appeared yellow in colour which suggests that the PLP cofactor was present as a ligand. Another apparent physical feature of the crystals was their needle-like shape, which was not improved upon several rounds of optimization screening. The crystals also appeared to be 'soft and bendy' potentially suggesting mis-alignment of unit cells within the crystal lattice. This feature may be due to the flexible and dynamic nature of the ACP domain and the disordered linker region which connects the ACP domain to the AOS

domain. The probability of disorder within the protein was analysed using regional order neutral network (RONN) algorithm,[175] which suggests residues 98-131 have a higher than 50% likelihood of being disorderd (Figure 40). These residues corresponded to the predicted linker region between the two domains. If this prediction is correct and this region is truly disordered then the likelihood of well diffracting crystals growing may be low.

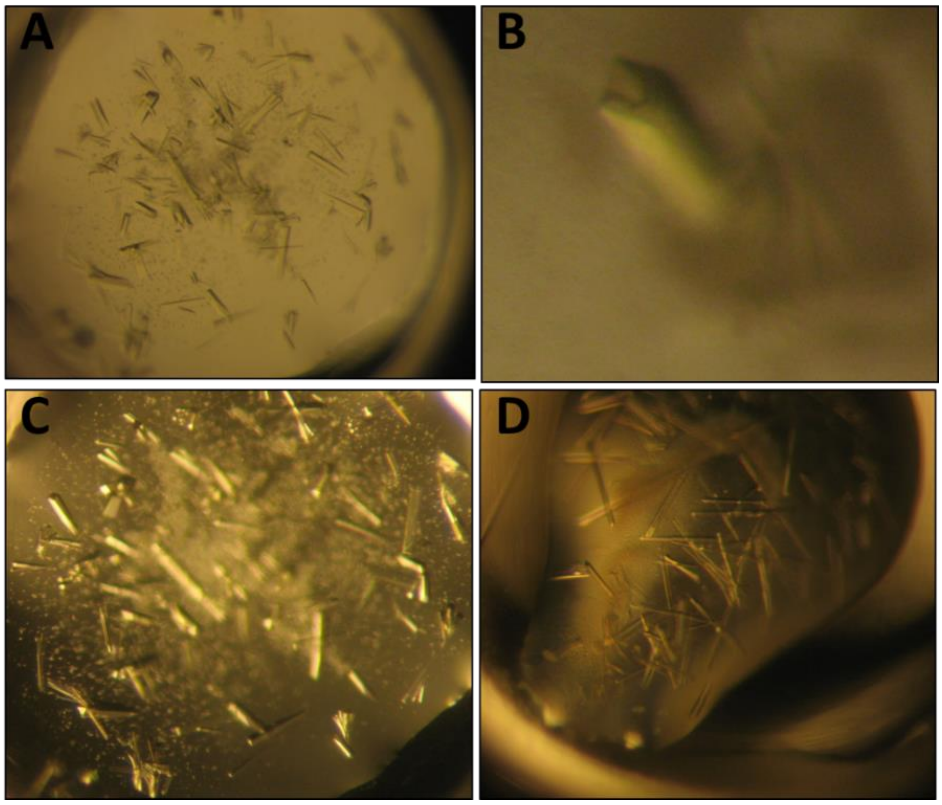


Figure 39: Crystals of TamD. Crystals were grown by vapour drop diffusion at 20 °C. The optimal growth conditions for crystal growth were 18.6 % PEG 3350 and 0.7 M $MgSO_4$ at a protein: precipitant ratio of 2:1.

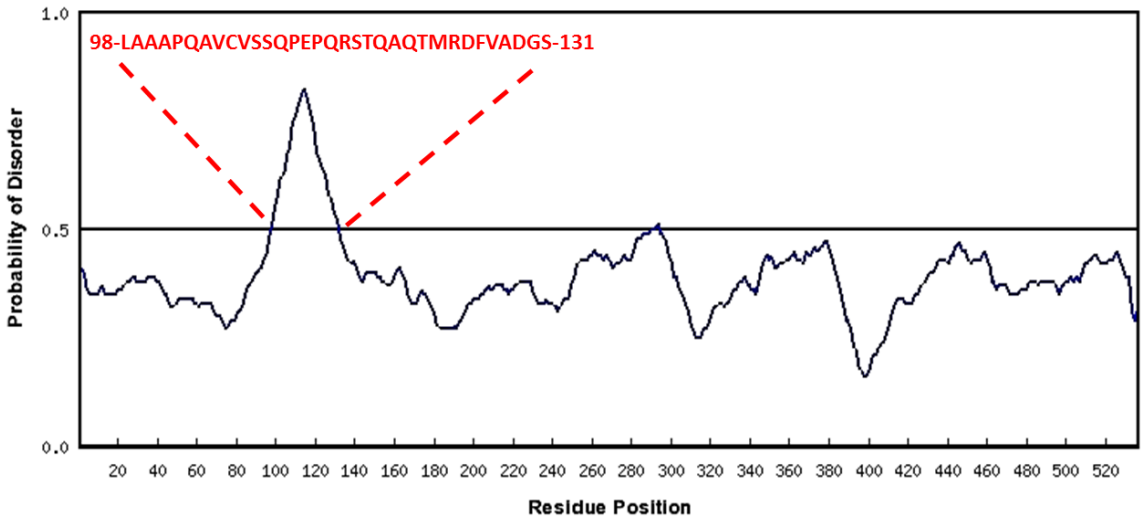


Figure 40: Probability of residue disorder in TamD. Residues 98-131 have the highest probability of disorder >0.5. Analysis was performed using the regional order neutral network (RONN) algorithm, which detects natively disordered regions in proteins.

3.3.5.2 Data Collection and Structural Solution

Despite a less than ideal morphology, the crystals diffracted weakly to a resolution of 4.98 Å, at beamline i24 at the DIAMOND synchrotron source, Didcot, Oxfordshire. The data collected were processed in an automated manner using xia2[176] and the statistics for these data are presented in Table 16. That the crystals diffracted only to this resolution may be partly due to a high solvent content and weakly packed lattice structure. Taking the output from xia2 a structure was solved by molecular replacement using the program Phaser[165] incorporated within CCP4[177] using a search model from *Sphingobacterium multivorum* SPT monomer (PDB code: 3A2B).[178] The Matthew's coefficient of 3.14 suggested that the expected number of molecules in the asymmetric unit was 4 in the $P4_1$ space group with a solvent content of 60.8 % (Figure A6). The solution from Phaser was refined in REFMAC and a concomitant decrease in R-factor and R-free was observed, suggesting the structure was solved (Figure A7). The partially refined model from

REFMAC was visualized in COOT.[167] No further refinement of the model was performed. The data collection and partial refinement statistics are summarised in Table 16.

Table 16: Data refinement and statistics for TamD. Statistics for the highest-resolution shell are shown in parentheses.

Wavelength (Å)	0.9686
Unit cell parameters (Å)	a=b=129.9, , c=174.1, alpha=beta=gamma=90^{O}
Space group	$P4_1$
Resolution range (Å)	43.3 – 4.98 (5.11 – 4.98)
I/σ(I)	8.4 (2.0)
R merge	0.116 (0.624)
Completeness (%)	98.2 (99.8)
Multiplicity	3.5 (3.8)
R factor	0.383
R free	0.448

3.3.5.3 Structural Analysis

The electron density obtained for TamD was insufficient to build a complete model of the structure. Unfortunately, the N-terminal domain was not visible and only traceable density for the 'SPT-like' AOS domain was detected (Figure 41). An analysis of the TamD primary sequence suggests that the protein is dimeric and belongs to the PLP-dependent aspartate aminotransferase (AAT) superfamily (fold I). The AOS domain dimer was visible in the low resolution crystal structure of TamD.

The low resolution data obtained for TamD at 4.98 Å, was also insufficient to support the level of details required to accurately describe the topology of the ACP domain in relation to the AOS domain of TamD. There are two postulated arrangements of the ACP domain relative to the AOS domain of the two monomers, head-to-head or head-to tail topology. In the head-to-head arrangement, ACP1 from monomer1 interacts with the

AOS1 from the same monomer (Figure 42A), whereas in the head-to-tail arrangement, ACP1 from monomer 1 interacts with AOS2 from monomer 2 (Figure 42B). A comparison of these two models with the low resolution crystal structure of a TamD-AOS dimer suggests that the more likely arrangement would be the head-to-tail topology (Figure 42C). This conclusion was drawn based on the position of the N- termini of each monomer (where the ACP domain would likely be found) relative to the likely location of the active site of the AOS domain (PLP-binding domain). Crystallographic analysis of TamD at a higher resolution should be able to provide more convincing evidence to substantiate this conclusion. Such data so far has been elusive, as the ACP domains were not visible from the electron density map of TamD.

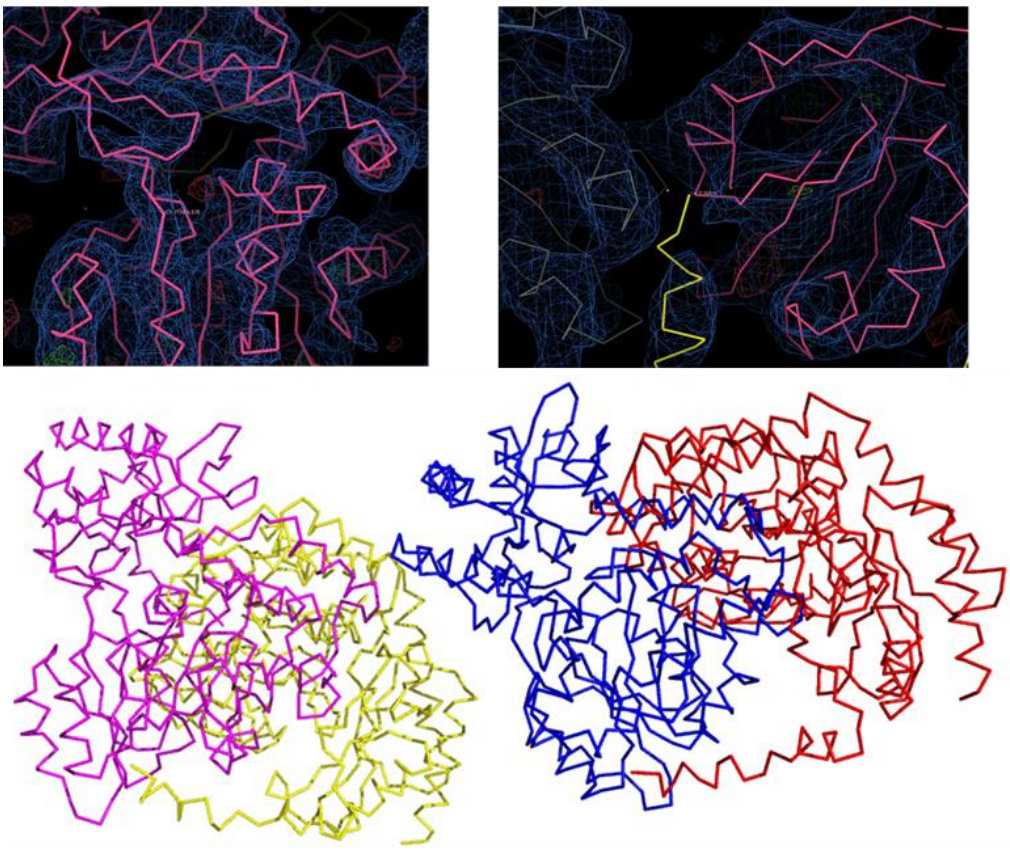


Figure 41: Crystal structure of TamD. The crystals diffracted to a resolution of 4.98 Å. Four molecules are present in the asymmetric unit (coloured in magenta, yellow, blue and red). Clear electron density exists only for the AOS domain as shown in sections of the electron density map. The N-terminal domain was not visible in the structure.

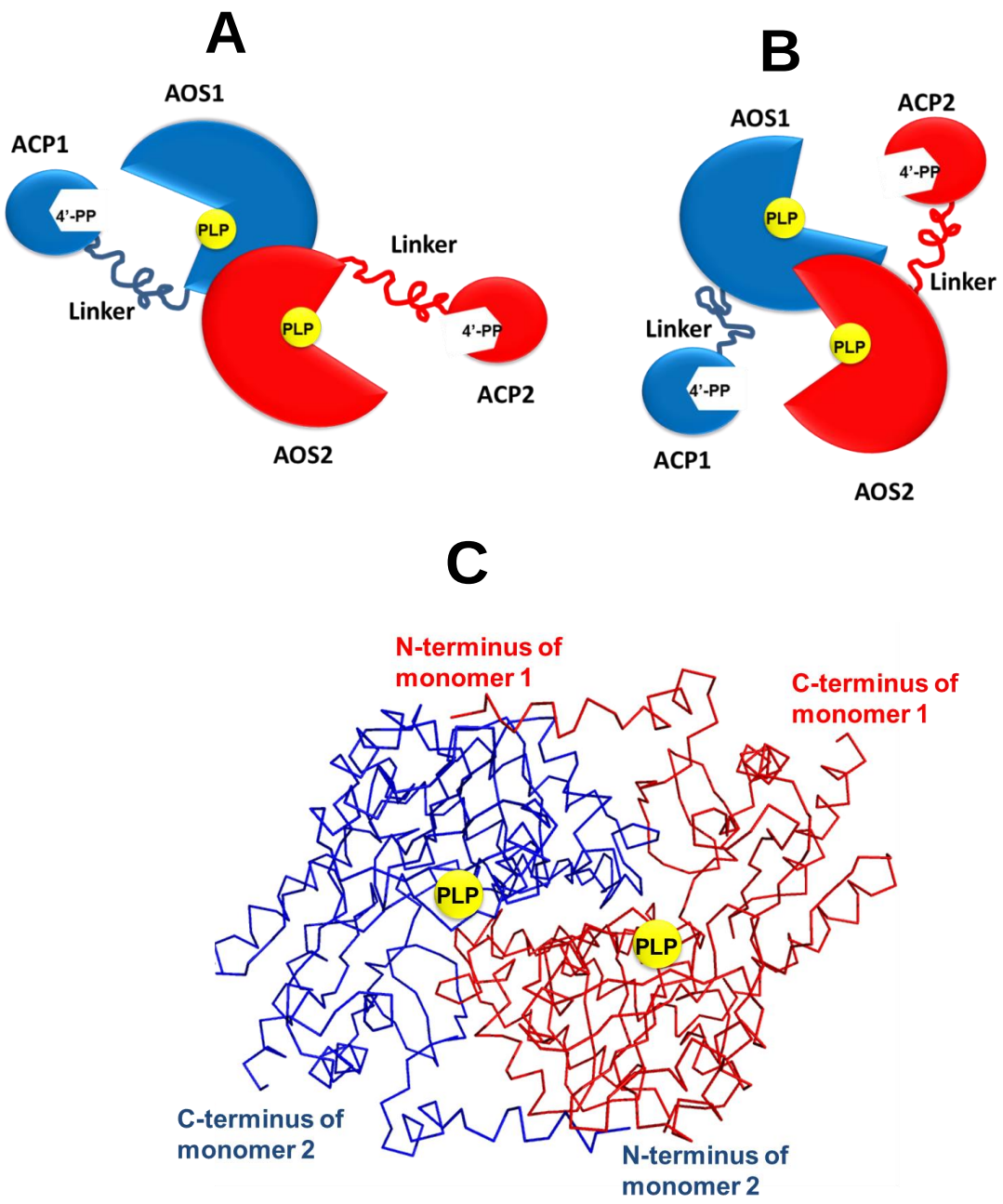


Figure 42: Schematic representation of the arrangment of the ACP and AOS domains of TamD. (A) In this head-to-head topology, the ACP domain of one monomer interacts with the AOS domain of the same monomer. (B) In the head-to-tail topology, the ACP domain of one monomer partners with the AOS domain of the second monomer. (C) The x-ray crystal structure likely suggests a head-to-tail arrangement based on the positions of the N and C termini of AOS monomers. ACP domain is absent from the structure.

3.4 Conclusion and Future Work

In this investigation, a novel AOS enzyme, TamD, from the tambjamine biosynthetic pathway was characterized using UV/vis spectroscopy and MS. This enzyme is predicted to catalyse the condensation between L-serine and a pyrrolyl-β-ketoacyl-S-ACP intermediate of the tambjamine biosynthetic pathway to give a bipyrrole product, HBM, in a PLP-dependent reaction. One interesting structural feature of the enzyme is the presence of an ACP domain at its N-terminus which is unsual for this family of enzymes. Extensive structural studies were carried out on the enzyme to determine the topological relationship between the ACP domain and the AOS domain. It is likely that the ACP domain of one monomer interacts with the AOS domain of the second monomer within the TamD dimer. However, this conclusion would be better supported with data from a crystal structure of a higher resolution than 4.98 Å.

In order to carry out further structural analysis on TamD, it is necessary to produce a construct where the residues of the ACP domain are well-ordered and in contact with the AOS domain. One approach in achieving a more ordered complex is to trap the acylated ACP domain by crosslinking it to the AOS domain. A mechanism-based ACP acyl substrate can be used to facilitate an active site-selective covalent crosslinking to a residue of the AOS domain. There are two possible histidine residues in the active site which are involved in PLP stabilization and may likely interact with the synthetic ACP substrate crosslinker, His273 and His349. A similar strategy was employed by the Burkhaart group, where *E. coli* AcpP was crosslinked to FabA, a fatty acid 3-hydroxyacyl-ACP dehydratase via a sulphonyl-3-alkyne-based probe.[79] Such a construct will more likely form a more ordered protein lattice that may improve the likelihood of diffraction at higher resolutions. Other approaches that may be useful in the biophysical characterisation of the protein-protein interaction between the ACP and the AOS domains include isothermal titration calorimetry (ITC) and surface plasmon resonance (SPR). Additionally, solving the structures of the ACP domain and the AOS domain separately would also provide useful information about their topological relationship.

Although the AOS domain was over produced and isolated in this study, attempts at crystallising the protein were unsuccessful. Another difficulty that was faced in this investigation was the phosphopantetheinylation of the ACP domain of TamD with the promiscuous PPTases, Sfp and AcpS. To overcome this problem, the dedicated PPTase from the Tam cluster, denoted as TamS was cloned. However, due to time constraints, the protein was not over produced in *E. coli*. Therefore in future studies, the isolation of TamS could be explored with the goal of using this enzyme to characterise TamD-ACP.

Chapter 4: Inhibition of Sortase A by *trans*-Chalcone

4.1 Introduction

4.1.1 Aims and Objectives

New approaches for the treatment and prevention of bacterial infections require an understanding of the molecular mechanisms by which protein targets interact with inhibitors. Sortases are attractive targets because of their role in the cellular location of multiple virulence factors in Gram Positive bacteria. Our collaborators on this project were interested in identifying flavonoids that could inhibit *S. mutans* sortase A, without the bactericidal effect on the organism. Before this study there was no biophysical data available on *S. mutans* SrtA. Therefore one of the aims of this study was to determine the three-dimensional structure of the enzyme and carry out its biochemical characterization. The inhibitory activity of *trans*-chalcone was also explored with the aim of gaining insights into the molecular basis of inhibition of the chalcone flavonoids using a combination of biophysical techniques: mass spectrometry, x-ray crystallography and molecular modelling. It was also of interest to investigate the impact of *trans*-chalcone on biofilm formation by *S. mutans* using a well established active attachment model.

4.1.2 *S. mutans* SrtA

Sortases are membrane-associated transpeptidases which are highly conserved in Gram positive bacteria.[113, 179] They are responsible for covalently attaching surface proteins involved in adherence to host cells, iron acquisition, invasion, signalling and pili formation to the bacterial cell wall.[113-115] In *S. mutans*, SrtA has been implicated in contributing to biofilm formation and the cariogenicity of the bacteria, where SrtA-deficient strains show reduced ability to colonise oral surfaces and cause infections.[129, 130, 132] These data suggest that SrtA makes a good target for the development of oral anti-infective therapies.[116] *S.aureus* SrtA was the first cysteine transpeptidase of this family to be identified in 1999 and its structure was later determined in 2001 by NMR

spectroscopy. Since then, several other structures of sortases from different Gram positive bacteria have been determined, some in complex with peptide substrate analogues and other small molecules (Table 17).[135] Sequence analysis has also shown that there is high structural homology among the *S. mutans* SrtA and other SrtA enzymes (Figure A8). The N-terminus region of sortases possesses hydrophobic residues which are membrane-associated and functions as a signal peptide for secretion as well as a stop transfer signal for membrane anchoring, while the C-terminal region, which is located across the cell membrane consists of the catalytic centre (Figure 43).[128, 134, 180]

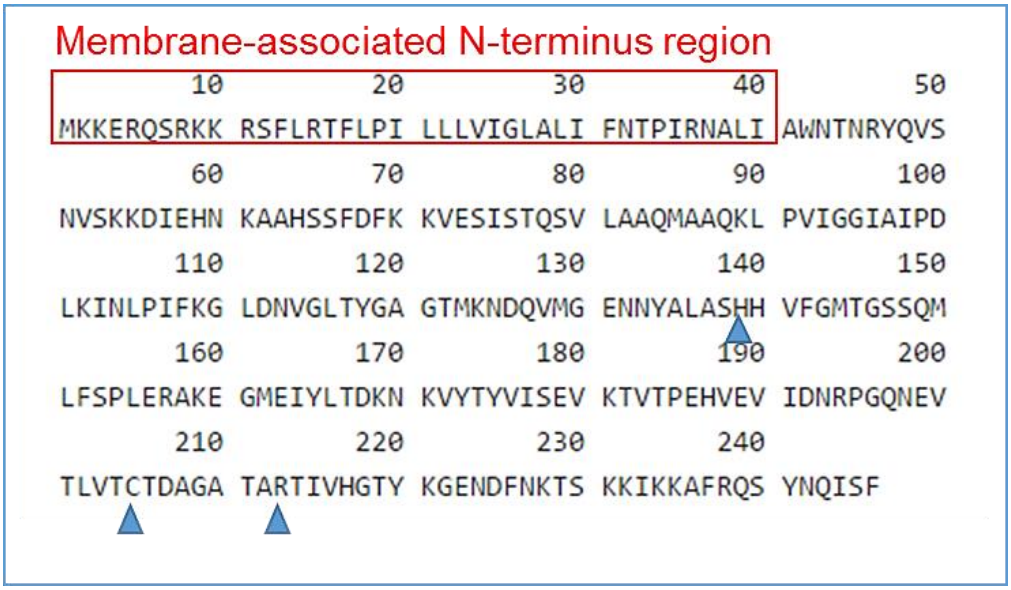


Figure 43: Annotated Primary Sequence of *S. mutans* SrtA. SmSrtA consists of 246 amino acids, with the first 40 hydrophobic residues membrane-associated (highlighted with a red box). The conserved catalytic triad (Cys-His-Arg) is located in the C-terminal region (highlighted with blue triangles).

The most recent report of natural products with inhibition activity against *S. mutans* SrtA identified two flavonoids, morin[154] and curcumin.[145, 155] Other flavonoids have also been reported in the literature to have inhibitory activities against other sortases.[153, 154] This is by no means surprising as flavonoids are known to display a broad spectrum of biological activities.[85] However, there is no information available on the mechanisms by which sortases are inhibited by these flavonoids. From these previous studies, *trans*-

chalcone, a widely occurring flavonoid precursor was selected as potential inhibitor of *S. mutans* SrtA and antibiofilm forming agent. This current study investigates the mechanism by which *trans*-chalcone interacts with *S. mutans* SrtA *in vitro* using a combination of techniques including mass spectrometry, enzyme kinetics and x-ray crystallography.

Table 17: Members of the sortase family whose structures have been determined to date. Structures of isoforms A, B and C have been determined by x-ray crystallography, some in complex with peptide ligands and small molecules.

Family	Name	Ligand	Organism	PDB Code
A	$SrtA_{\Delta 59}$	none	*S. aureus*	1IJA
A	$SrtA_{\Delta 59}$	LPETG Peptide	*S. aureus*	1T2W
A	Sp-$SrtA_{\Delta 81}$	none	*S.pyogenes*	3FN5
A	$SrtA_{\Delta 59}$	(PHQ)LPA(B27) Peptide	*S. aureus*	2KID
A	$SrtAH139A_{\Delta N40}$	none	*S.mutans*	4TQX
B	Sa-SrtB	none	*S. aureus*	1NG5
B	Sa-SrtB	MTSET	*S. aureus*	1QWZ
B	Sa-SrtB	E-64	*S. aureus*	1QX6
B	Sa-SrtB	Gly_3	*S. aureus*	1QXA
B	Ba-*SrtB*	AAEK1	*B.anthracis*	2OQW
B	Ba-*SrtB*	AAEK2	*B.anthracis*	2OQZ
B	Sa-SrtB	(CBZ)NPQ(B27) Peptide	*S. aureus*	4LFD
C	SrtC-1	none	*S.pneumoniae*	2W1J
C	SrtC-2	MES	*S.pneumoniae*	3G66
C	SrtC-3	none	*S.pneumoniae*	2W1K
C	SrtC-1 H131D	none	*S.pneumoniae*	2WTS
C	SrtC-1	none	*S. agalactiae*	3O0P
C	SrtC-1	none	*S. suis*	3RE9
C	Srt-C-1	none	*A. oris*	2XWG

4.2 Results and Discussion

4.2.1 Cloning and Expression

4.2.1.1 SmSrtA-N40 (SmSrtA)

The sequence of the wild-type SrtA from *S. mutans* G-S5 strain (SmSrtA WT) has a length of 741 bp which encodes a full length protein of 246 amino acids.[128] Homology searches in the NCBI data base revealed several sortase homologues in Gram positive bacteria with highly conserved active site residues. The gene sequence for SmSrtA WT was downloaded from the GeneBank database (accession number AB089932) and used as a template for designing a synthetic gene (GenScript), with the DNA sequence modified to remove the first 120 bp and insert the restriction endonuclease sites *Nco1* (5') and *Xho1* (3'). This allowed for the insertion of the synthetic gene into a pET28a vector for protein expression. Overexpression of pET28a/SmSrtA-N40 produced an untagged recombinant protein with an N-terminal truncation of the first 40 amino acids (SmSrtA). This specific truncation was chosen to increase the likelihood of the recombinant enzyme being soluble upon expression and purification in *E.coli*, as the N-terminal domain is known to possess hydrophobic residues which are membrane-associated. [128, 180 134] In previous studies similar constructs of other SrtA homologues were cloned, without the hydrophobic N-terminal region to produce soluble and active protein. Optimal expression was obtained from *E. coli* BL21 (DE3) cells grown at 20 ^{o}C and induced for 5 hours with 0.1 mM IPTG.

4.2.1.2 SmSrtA Mutants (H139A and G94A)

Two other constructs of SmSrtA were made with alanine substitutions at the G94 and H139 positions, so as to improve protein stability. The G94A mutation was chosen based on the hypothesis that the WT enzyme was being proteolysed around residues 90-95 (LPVIGG). This 6-amino acid sequence bears similarities to the 5-amino acid cleavage site (LPXTG) that is recognized by sortase enzymes in their peptide substrates. The H139A mutant is expected to produce protein that is inactive due to abolished sortase

activity, which should in turn improve protein stability.[181] The pET28a plasmids encoding the mutant enzymes were purified, sequenced for confirmation and then used to transform *E. coli* BL21 (DE3) for protein expression.

4.2.1.3 SmSrtA-N88 H139A

A third construct with a further truncation of 48 additional amino acids from the N-terminus of SmSrtA was produced from the gene encoding for the H139A enzyme. This construct was produced in an effort to investigate the role of the N-terminal domain in the reactivity of the enzyme towards covalent modifying small molecules. The gene was amplified with the restriction endonuclease sites *Nco1* (5') and *Xho1* (3') appended. This allowed for the insertion of the gene into a pET28a expression plasmid. Overexpression of this construct produced an untagged recombinant protein with an N-terminal truncation of the first 88 amino acids.

4.3.1 Protein Purification and Analysis

4.3.1.1 SmSrtA

SmSrtA was isolated from cell free extract (CFE) by cation exchange chromatography which was possible due to the protein's high isoelectric point (pI) of 9.01. The protein eluted from a MonoS 5/50 column with increasing concentration of NaCl in 45-50% of Buffer D (Figure 44A). The fractions containing SmSrtA were pooled and further purified by SEC on a Superdex S200 16/60 column to give a symmetric peak corresponding to the monomeric mass of SmSrtA, approximately 23 kDa (Figure 44B). SDS/PAGE analysis of the purified enzyme showed a band corresponding to the enzyme running at ~28 kDa. SrtA enzymes usually run higher on SDS/PAGE gels due to the positive nature of these proteins. SDS-PAGE analysis of WT SmSrtA also revealed auto-proteolysis of the enzyme after purification.

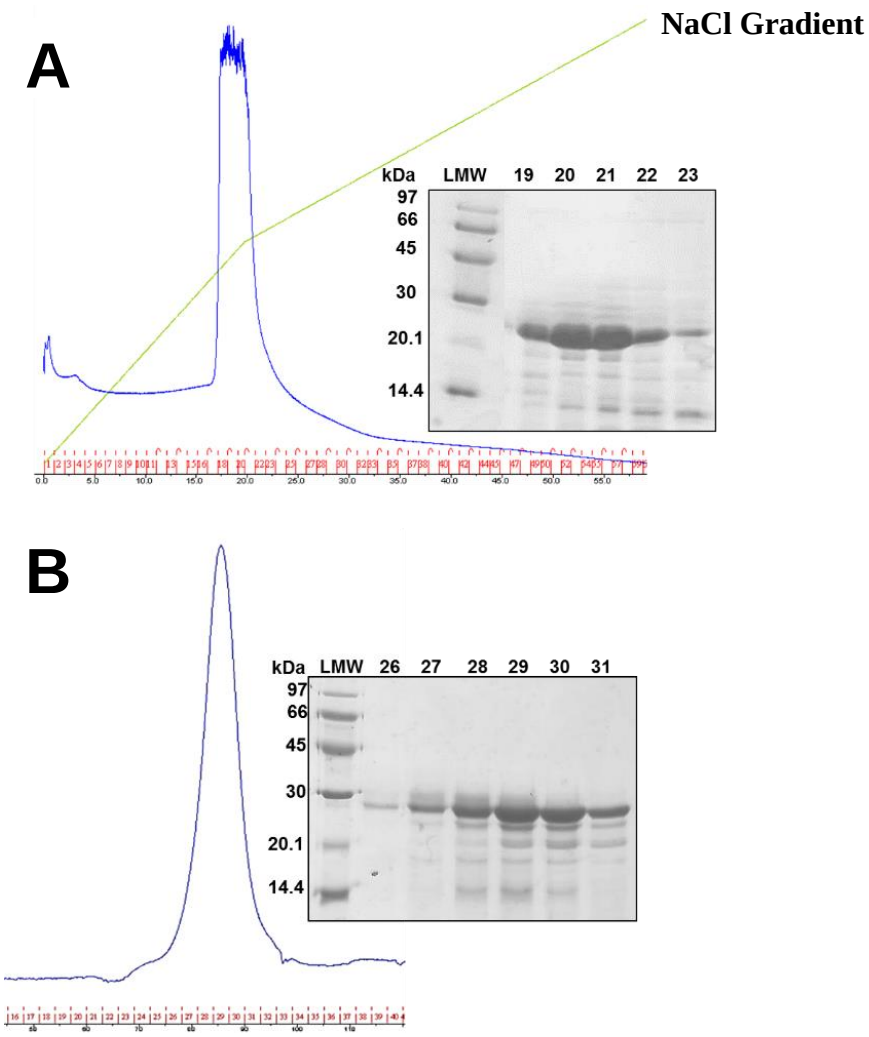


Figure 44: Purification of recombinant SmSrtA from *E. coli*. (A) Elution profile of enzyme by cation exchange chromatography and SDS-PAGE analysis of purified enzyme (inset). (B) Elution profile of enzyme by size exclusion chromatography on superdex S200 16/60 and SDS-PAGE analysis of purified enzyme (*inset*).

4.3.1.2 G94A and H139A

Both mutant enzymes, G94A and H139A, were purified using the protocol described above in section 4.3.1.1 for the WT enzyme via cation exchange chromatography followed by gel filtration. The G94A mutant showed no significant reduction of the

enzyme's degradation, which suggests that the cleavage site may not be between Ile93 and Gly94. Large aggregates were also observed in the elution profile on purification (Figure 45A). On the other hand, the H139A mutant which is expected to be catalytically inactive showed no signs of auto-proteolysis upon purification. This suggests that protease activity was abolished, which resulted in increased stability of the protein on purification. The H139A mutant eluted in the same volume of buffer as the WT enzyme (~85 mL) (Figure 45B). No degradation products were observed from SDS/PAGE analysis, which also suggests improved stability over the WT enzyme. This observation also confirms the essential role of the histidine residue in catalytic activity.

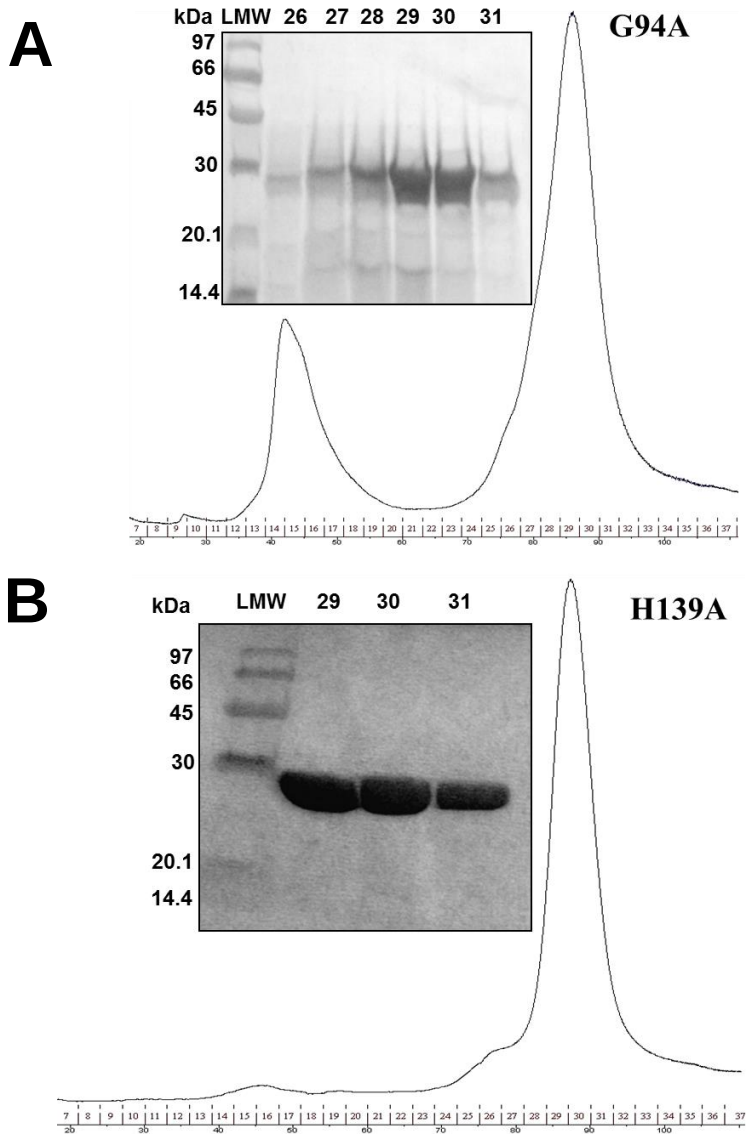


Figure 45: Purification of *S. mutans* SrtA mutants. (A) Elution profile of SmSrtA G94A mutant showing a symmetric peak for the protein eluted over 60 mL of buffer off a Superdex S200 16/60. Fractions 26-31 were analysed by SDS-PAGE to give a monomeric mass of ~23 kDa (inset). (B) Elution profile of SmSrtA H139A showing a similar elution pattern as that of WT SmSrtA, under similar purification conditions. Fractions 29-31 were analysed by SDS-PAGE to give a monomeric mass of ~23 kDa (*inset*).

4.3.1.3 SmSrtA-N88

SmSrtA-N88 was purified using the same protocol described above in section 4.3.1.1 for the WT enzyme via cation exchange chromatography followed by gel filtration resulting in soluble monomeric protein in solution. However, the protein degraded readily upon purification. This could be due to the fact that the construct was unstable without the additional 88 amino acids at its N-terminus, which may contribute to enzyme stability.

4.3.2 Enzyme Activity

All enzyme activity assays were performed using the SmSrtA construct.

4.3.2.1 Transpeptidation Reaction

The purified enzyme SmSrtA was first analysed for its ability to cleave peptide bonds between a threonine and a glycine residue and form a new amide bond between two glycine residues. This was done by incubating the purified enzyme (5 µM) with the peptide substrate analogue, dabcyl-QALPETGEE-edans (25 µM) for 16 h in a reaction buffer containing 20 mM MES.NaOH, 150 mM NaCl and 5 mM DTT.[139] The substrate peptide analogue has the LPXTG motif (where x is glutamic acid) tethered to a fluorophore (Edans) and a fluorescence quencher (dabcyl) at either ends of the short peptide. The transpeptidation reaction mixture was incubated in the presence of NH_2-Gly_3 (5 mM) which acted as the nucleophile substitute for the natural substrate, lipid II which bears a pentaglycine side chain. The reaction mixtures were then analysed by LC-ESI-MS. In the absence of the NH_2-Gly_3, SmSrtA cleaved the peptide substrate between the threonine and glycine residues to generated two fragments, dabcyl-QALPET and GEE-edans, with observed average masses of 909.9 Da (calculated mass, 909.0 Da) and 581.2 Da (calculated mass, 581.6 Da) respectively (Figure 46A). In the presence of NH_2-Gly_3, SmSrtA generated a new peptide, dabcyl-QALPET-Gly_3 of mass of 1081.0 Da (calculated mass, 1080.2 Da) (Figure 46B). The masses of the peptide fragments are consistent with the calculated values for each peptide fragment and previous values

reported in the literature.[139] These data suggest that the truncated enzyme has the ability of recognizing the synthetic substrates to catalyse the transpeptidation reaction.

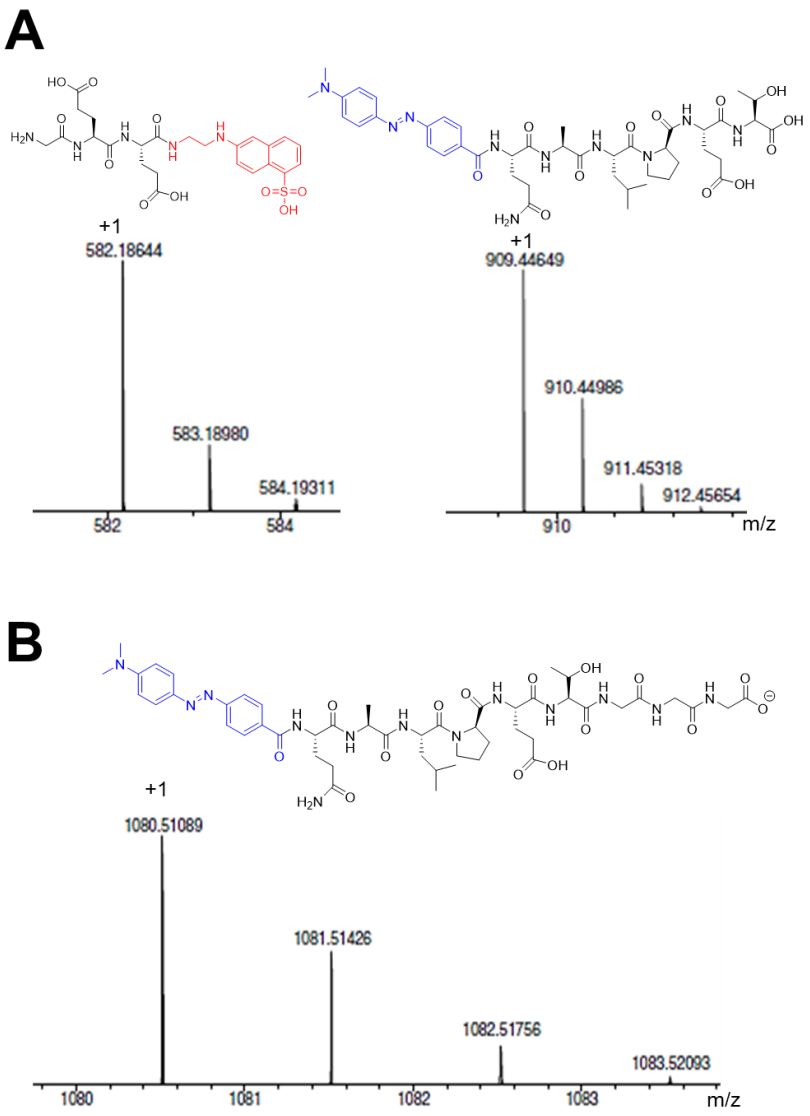


Figure 46: Mass spectrometry analysis of sortase activity of SmSrt. (A) Mass spectrum of the cleavage reaction of dabcyl-QALPETGEE-edans, showing the charge distribution of a singly charged ion for the peptide fragments of average masses 582.18 Da and 909.9 Da. (B) Charge distribution for the transpeptidation product formed btween dabcyl-QALPET and triglycine with average mass 1081.0 Da.

4.3.2.2 pH Optimisation of Sortase Activity

The optimal pH of SmSrtA activity was determined using the FRET assay described previously. The assay was performed at different pH values ranging from pH 4.0-9.0. The purified enzyme (5 µM) was incubated in buffer L (50 mM acetic acid, 50 mM MES, 100 mM Tris) adjusted to pH 4.0-5.5 or buffer M (50 mM MES, 100 mM Tris, 50 mM CAPS) adjusted to pH 5.5-9.0 in 0.5 pH unit increments. SmSrtA was found to be soluble and active over the pH range that was tested. Interestingly, it was found that SmSrtA had a greater activity at a lower pH and cleaved its peptide substrate at an optimum pH of 4.5 (Figure 47). pH 6.5 was used in subsequent studies since the enzyme degraded more rapidly in buffers of pH ≤ 5.5 after 18 h at 4 ^{O}C.

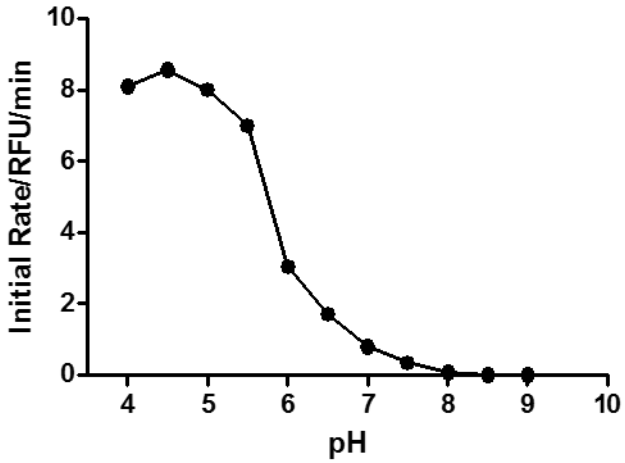


Figure 47: pH Profile of SmSrtA Activity. The activity of SmSrtA was analysed using the FRET assay at pH values ranging from 4-9. The enyme had maximum activity at pH 4.5. pH 4.0-5.5 was obtained in buffer L and pH 5.5-9.0 was obtained in buffer M.

4.3.3 Kinetic Analysis of SmSrtA

4.3.3.1 Calibration Curve for FRET Assay

The established FRET assay was used to measure the rate at which SmSrtA cleaved the fluorogenic peptide substrate, Dabcyl-QALPETGEE-Edans. Standards containing an equimolar mixture of free Dabcyl and Edans ranging from 0-10 µM were used to prepare a standard plot of relative fluorescence intensity vs concentration of Edans. This analysis was repeated using standards of only free Edans. The standard curves for both solutions gave straight lines with correlation coefficients of r^2 = 0.9964 and 0.9955 respectively and gradients of 19.03 ± 0.261 and 21.88 ± 0.336 (Figure 48). In the presence of high concentrations of Dabcyl, the fluorescence signal of Edans is lowered. However, this effect was minimal at lower concentrations of both molecules (0-3 µM) as less molecules of Dabcyl are present in solution to quench the fluorescence signal of Edans. The standard curves showed good agreement at lower concentrations, which fall within the range of concentration values that would be analysed in this study. Based on the Edans/Dabcyl standard plot, the net gain was measured to be equivalent to 23 F.U./µmol upon conversion of the fluorogenic peptide substrate to the Edans peptide product (GEE-Edans).

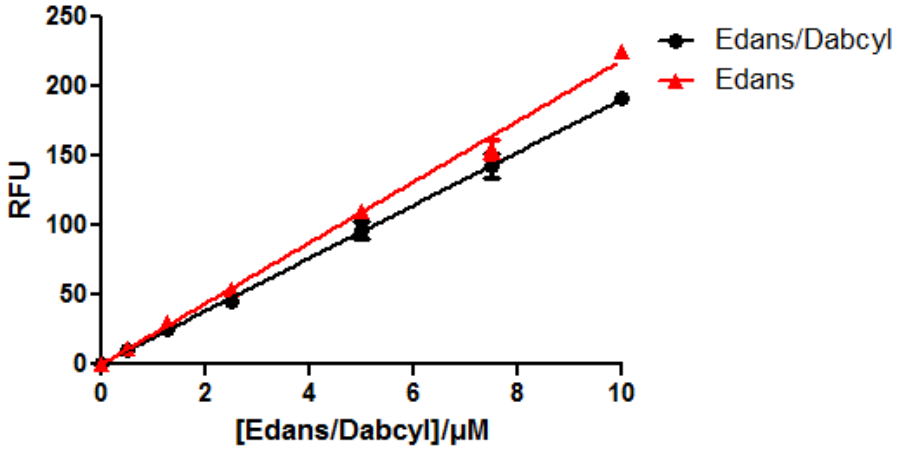


Figure 48: Standard curves for Edans and Dabcyl/Edans. Standards contained an equimolar mixture of free Dabcyl and Edans ranging from 0-10 µM. The standard curves for both solutions gave straight lines with correlation coefficients of r^2 = 0.9964 and 0.9955 respectively and gradients of 19.03 ± 0.261 and 21.88 ± 0.336.

4.3.3.2 K_m and V_{max}

The established FRET-based assay was used to measure the rate at which SmSrtA cleaved the fluorogenic peptide substrate. The purified WT enzyme (5 µM) was incubated with different concentrations of the peptide substrate ranging from (0-212 µM) and the increase in fluorescence was measured over 3 h at 1 min intervals at a wavelength of 485 nm.[182] To determine the rates of peptide cleavage (µM/min), the increase in fluorescence intensity (RFU) was plotted over time (min) using data points from only the linear portion of the hydrolysis curves for each concentration. The slope from these plots were converted to molar velocities using the net gain of 23 FU/µmol as obtained from the calibration curve of an equimolar mixture of free *dabcyl* and *edans* standards (Figure 48). The kinetic constants K_m and V_{max} for the hydrolysis reaction by SmSrtA were calculated from the Michaelis-Menten plot (Figure 49). SmSrtA exhibited K_m and V_{max} values of 90.4 ± 4.7 µM and $4.46 \times 10^{-4} \pm 0.11 \times 10^{-4}$ µM/s respectively.

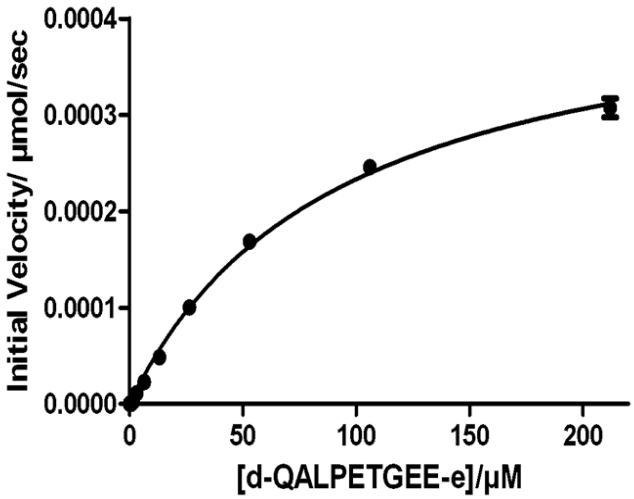


Figure 49: Kinetic analysis of SmSrtA using the FRET assay. Initial velocity was determined from the linear portions of the curves obtained from monitoring the cleavage of the peptide substrate. These were converted to µmol/sec using the net gain of 23 FU/mol. Curve was fitted to the Michaelis-Menten equation to obtain Km and Vmax values. All reactions were performed in triplicate.

4.3.3.3 Inhibition of SmSrtA by *trans*-Chalcone

Purified SmSrtA (5 µM) was incubated with *trans*-chalcone (0-100 µM) for 16 h. The activity of the pre-treated enzyme was measured using the FRET assay described previously by monitoring the cleavage of the Dabcyl-QALPETGEE-Edans peptide. The % activities at the different inhibitor concentrations were then plotted against the corresponding concentration of *trans*-chalcone and the IC_{50} of the inhibitor determined. The inhibitory activity of *trans*-chalcone is defined as the concentration giving 50% inhibition of SmSrtA activity, relative to 1% DMSO which was considered as the negative control. Curcumin and morin served as positive controls in the experiment. *Trans*-chalcone inhibited *S. mutans* SrtA with an IC_{50} value of 5.0 ± 0.6 µM (Figure 51). This low value suggests that *trans*-chalcone is a tight binding inhibitor of the enzyme. The IC_{50} of *trans*-chalcone was also lower than the values observed for other flavonoids reported in the literature, morin (IC_{50}= 27.2 ± 2.6 µM)[154] and curcumin (IC_{50}= 10.2 ± 0.7

μM)[155]. In this study, the IC_{50} of morin and curcumin were determined as 36.3 ± 1.5 μM and 14.7 ± 1.1 μM respectively (Figure 50).

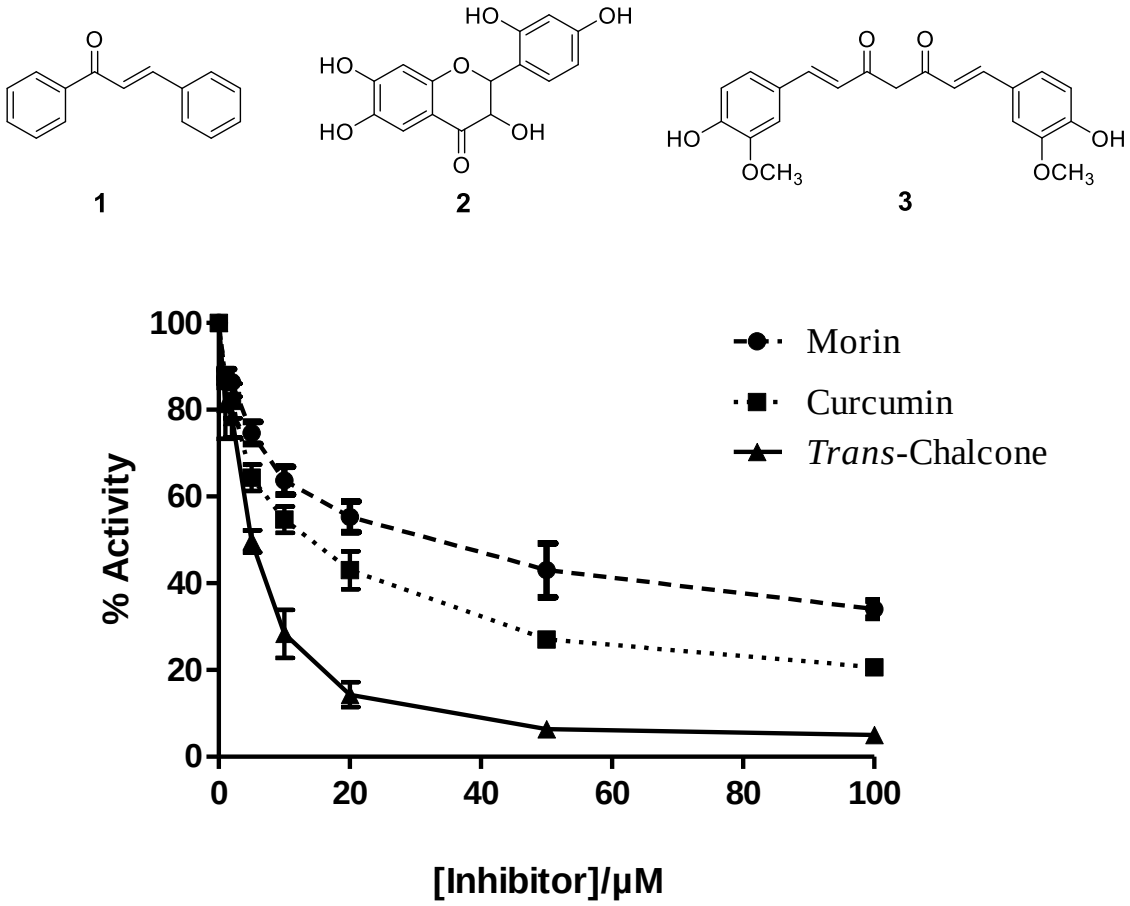


Figure 50: Concentration-dependent inhibition of SmSrtA *trans*-chalcone, morin and curcumin. Inhibition of SmSrtA by *trans*-chalcone (**1**) after overnight incubation with varying concentrations of the inhibitor at 4°C. Inhibition activity was monitored in the FRET assay with 25 μM of the *dabcyl*-QALPETGEE-*edans* peptide substrate. The IC_{50} of *trans*-chalcone was determined as 5.0 ± 0.6 μM. Morin (**2**) and curcumin (**3**) were used a positive controls. Enzyme activity was assessed using the FRET assay and all reactions were performed in triplicate.

4.3.3.4 Rate of inhibition

The rate of inhibition was measured in a time-dependent study by incubating 5 μM of the purified SmSrtA with excess *trans*-chalcone for varying time points (1, 2, 4, 8 and 16 h). Analysis of the enzyme activity revealed that the inhibition of SmSrtA by *trans*-chalcone was very slow and it took up to 16 h to reduce enzyme activity by 90% (Figure 51). Enzyme activity was not restored upon extensive dialysis of the enzyme against 20 mM MES.NaOH, 250 mM NaCl, pH 6.0. The data from the concentration- and time-dependent studies suggested that *trans*-chalcone may be a slow tight binding inhibitor of SmSrtA. This nature with which the inhibitor interacted with SmSrtA led to the postulation that a covalent adduct was formed between the inhibitor and SmSrtA.

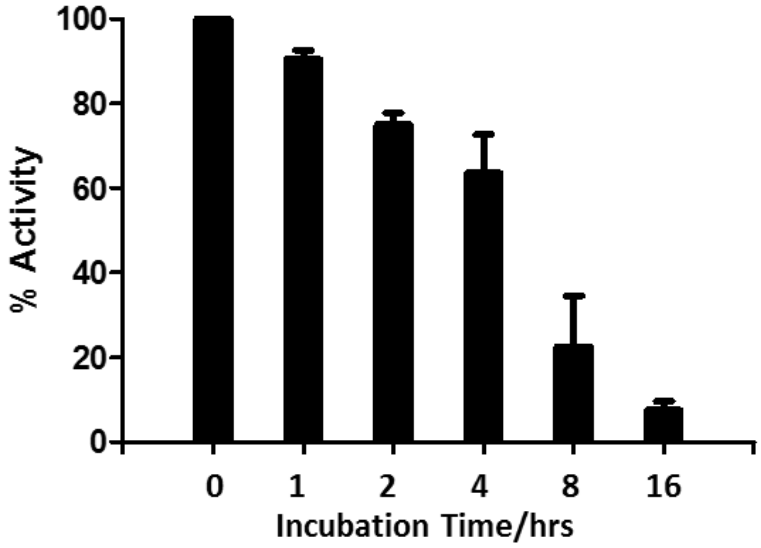


Figure 51: Time-dependent inhibition of SmSrtA by *trans*-chalcone. Inhibition of SmSrtA by 100 μM of *trans*-chalcone monitored at varying time points (1, 2, 4, 8 and 16 h). Enzyme activity was assessed using the FRET assay and all reactions were performed in triplicate.

4.3.3 Mass Spectrometry Analysis

All MS analysis in this chapter were done by Dr. David Clarke at the School of Chemistry, The University of Edinburgh.

4.3.4.1 ESI-MS of SmSrtA

The mass of SmSrtA was confirmed by ESI-MS resulting in an average mass ($[M+H]^+$) of 22, 767 Da in accordance with the molecular mass deduced from the SmSrtA gene sequence minus the N-terminal methionine which is post-translationally cleaved during expression in *E. coli* (Figure 52). This mass corresponds to the monomeric form of the protein.

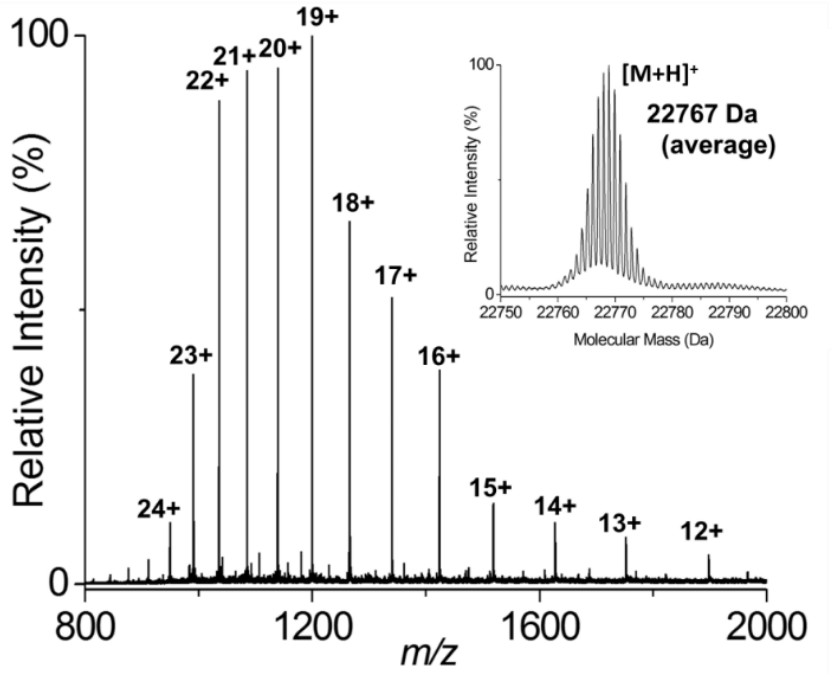


Figure 52: Mass spectrum of truncated SmSrtA. The main figure shows the charge state distribution obtained for SmSrtA. This corresponds to a deconvoluted molecular mass of 22, 767 Da, ($[M+H]^+$) see insert, which was consistent with the predicted theoretical mass based on the amino acid sequence of the protein (theoretical $[M+H]^+$, 22, 767 Da).

4.3.4.2 SmSrtA Inhibition by *Trans*-chalcone

In order to test the hypothesis that *trans*-chalcone was covalently modifying SmSrtA, MS was used to analyse the samples from the concentration- and time- dependent reactions described previously. The mass spectra from the concentration-dependent studies revealed a peak of 22, 767 Da for the mass of SmSrtA and an additional peak at 22, 975 Da (Figure 53A). This new peak represents a mass increase of +208 Da, which corresponds to the formation of a covalent adduct between *trans*-chalcone (208 Da) and SmSrtA (22, 767 Da). Further confirmation of the new peak resulting from the reaction of the enzyme with *trans*-chalcone was obtained by monitoring the change in the charge distribution of SmSrtA and the covalent adduct (Figure 53B).

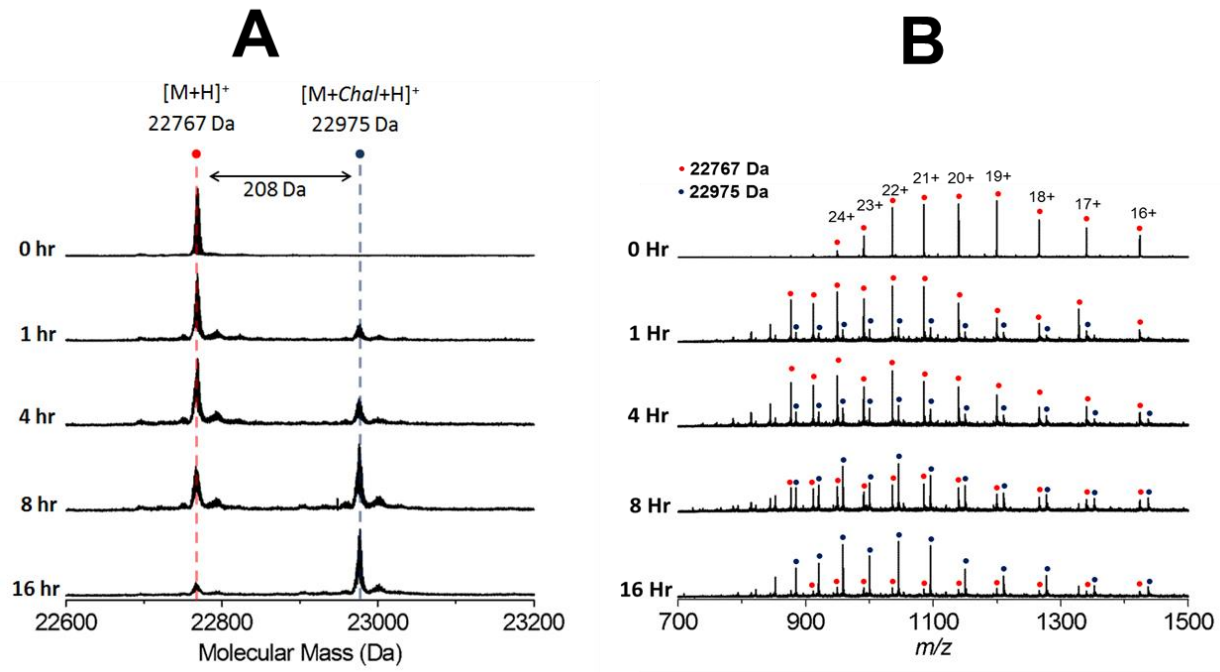


Figure 53: Mass spectrometry analysis of the time-dependent inhibition of SmSrtA by *trans*-chalcone. (A) Deconvoluted mass spectrum of *trans*-chalcone (208 Da) reacted with SmSrtA (22, 767 da) to form a covalent adduct (22, 975 Da) in a time-dependent manner after 16 h of incubation at 4 ^{O}C. (B) Changes in charge distribution over time as SmSrtA reacts with *trans*-chalcone to produce a covalent adduct. The distribution

highlighted with red dots is for SrtA (22, 767 Da) and the distribution for the adduct is highlighted with blue dots (22, 975 Da).

Adduct formation was observed to be dependent on the concentration of the inhibitor as the species of mass 22, 975 Da increased in abundance as higher concentrations of *trans*-chalcone were incubated with the enzyme. A similar result was observed for the samples analysed from the time-dependent study, where the new peak at 22, 975 Da also increased in abundance for samples incubated for longer periods (Figure 54).

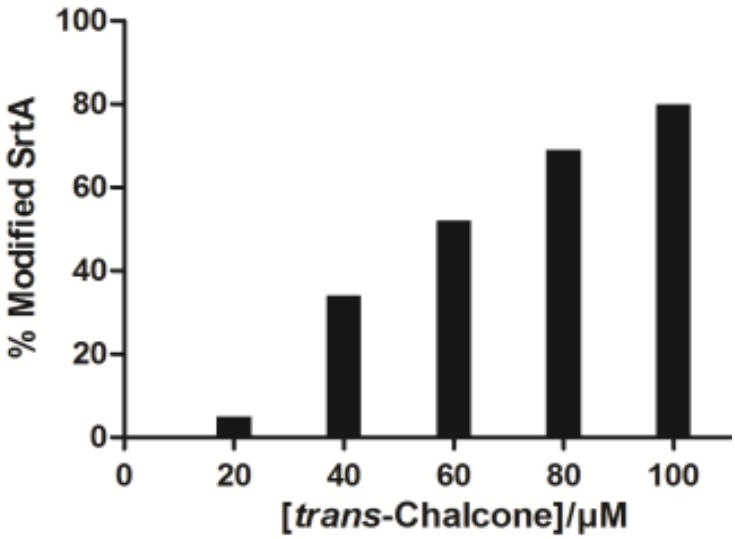


Figure 54: Mass spectrometry analysis of the concentration-dependent inhibition of SmSrtA by *trans*-chalcone. The % of modified SmSrtA increased when the concentration of *trans*-chalcone was increased in the reaction mixture. ~ 90% modification of SmSrtA by the inhibitor was observed on incubation of the enzyme with 100 µM of *trans*-chalcone for 16 h.

4.3.4.3 H139A SrtA-Chalcone Adduct

Despite being inactive in the transpeptidase assay, the SmSrtAH139A mutant reacted with *trans*-chalcone in the same manner as the WT SmSrtA, forming a covalent adduct of mass increase of 208 Da (Figure 55). This suggests that the H139A residue was not required for the modification of SrtA by *trans*-chalcone.

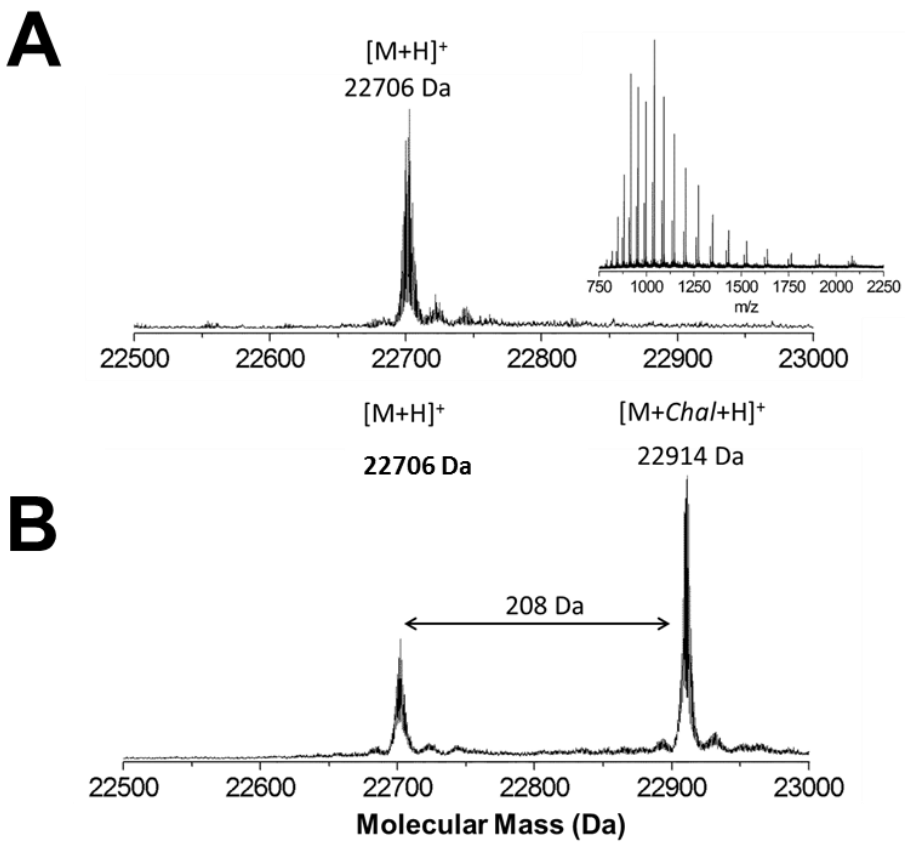


Figure 55: Mass spectra of H139A mutant before and after incubation with *trans*-chalcone. (A) Deconvoluted mass and chare distribution for SmSrtAH139A corresponding to the predicted monomeric mass of 22, 705 Da. (B) The mutant enzyme was incubated overnight with *trans*-chalcone giving a modified species with mass of 22, 914 Da (M+Chal+H^+).

4.3.4.4 Top-down Fragmentation

The most reactive residue in the primary sequence of SmSrtA that would likely undergo modification by *trans*-chalcone is expected to be the active site residue Cys205. To investigate this hypothesis, top-down fragmentation MS experiments were performed. In these experiments, SmSrtA (5µM) was incubated with excess *trans*-chalcone (100 µM) for 16 h, followed by ESI ionisation of the sample. The +24 charge state of the protein had an average mass of 22, 975 Da which suggests that majority of the enzyme was

modified by *trans*-chalcone (Figure 56A). This isolated charge state of the modified enzyme was further fragmented by collision induced dissociation (CID) and electron collision dissociation (ECD) sources (Figure 56B). The product ions from the fragmentation experiments were analysed using the Prosight search software against the primary sequence for the truncated SmSrtA. The combined product ion maps from both the ECD and the CID experiments had very good correlation of fragmentation (Figure 56C). The $\boldsymbol{y_{42}^{5+}}$ ($[C_{207}H_{333}N_{60}O_{64}S_1(C_{15}H_{12}O)]^{5+}$) and $\boldsymbol{y_{43}^{6+}}$ ($[C_{211}H_{341}N_{61}O_{66}S_1(C_{15}H_{12}O)]^{6+}$) fragment ions from the CID experiment were used for the assignment the site of modification as Cys205 (Figure A9).

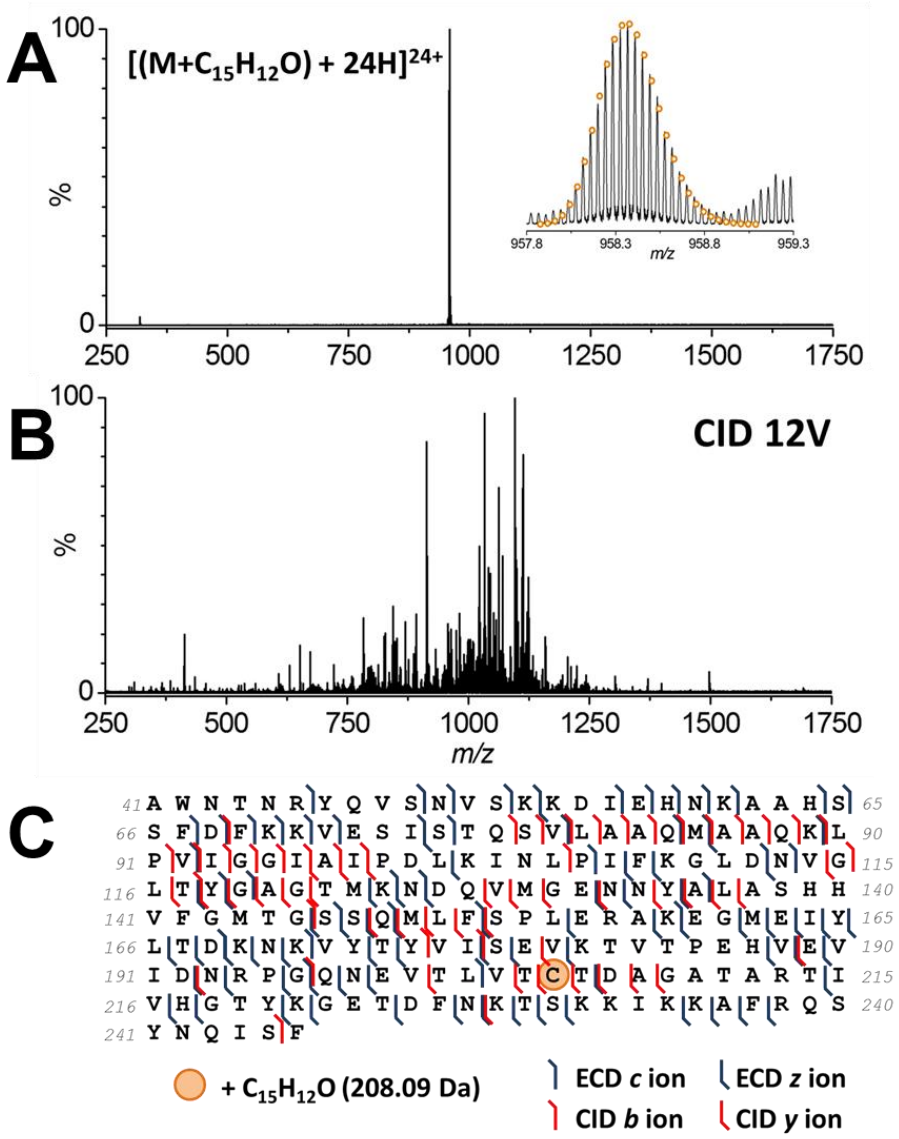


Figure 56: Mapping the location of the *trans*-chalcone modification in SmSrtA by top down MS. (A) Typical mass spectrum of a single isolated charge state of chalcone-modified sortase. Here the 24+ charge state is shown. *Inset*, Isotope distribution of the 24+ species. The ion has an isotope pattern consistent with the addition of chalcone ($+C_{15}H_{12}O$). The simulated isotope pattern is displayed in orange ($[C_{1023}H_{1634}N_{276}O_{311}S_7]^{24+}$). (B) Typical top-down mass spectrum. Here the CID spectrum of the 24+ species is shown. (C) Combined top down fragment map showing all of the assigned product ions from all fragmentation experiments. Fragments resulting from CID are displayed in red, fragments resulting from ECD are displayed in blue. Fragmentation allowed the assignment of Cys205 as the residue modified by chalcone.

4.3.5 Reaction Mechanism

4.3.5.1 *Trans*-Chalcone and SmSrtA

The proposed mechanism for the reaction of *trans*-chalcone with SmSrtA is via a Michael addition. A Michael addition reaction is characterized as a 1,4 conjugation addition which results when a strong nucleophile attacks the β-carbon of an α,β-unsaturated carbonyl to produce an adduct (Figure 57). The chalcone flavonoids have a characteristic α,β-unsaturated carbonyl system between two aromatic rings which can readily react with strong nucleophiles to produce Michael adducts. In the case of SmSrtA, the active site consists of a reactive thiol, Cys205 which serves as a strong nucleophile to carry out an attack on the β-carbon of *trans*-chalcone yielding a chalcone-SrtA adduct (Figure 57). This proposition was further strengthened by analyzing reactions of SmSrtA with dihydrochalcone (average mass: 210 Da) (Figure 58A), the saturated analogue of *trans*-chalcone, by FRET assay and MS. Dihydrochalcone showed no inhibition of enzyme activity (Figure 58B), likewise no additional peak of mass increase of +210 Da was observed in mass spectrum which would correspond to SrtA-dihydrochalcone adduct (Figure 58C). These observations therefore suggest that the α,β-unsaturated system of *trans*-chalcone is required for reaction with SmSrtA and supports the proposed Michael addition mechanism.

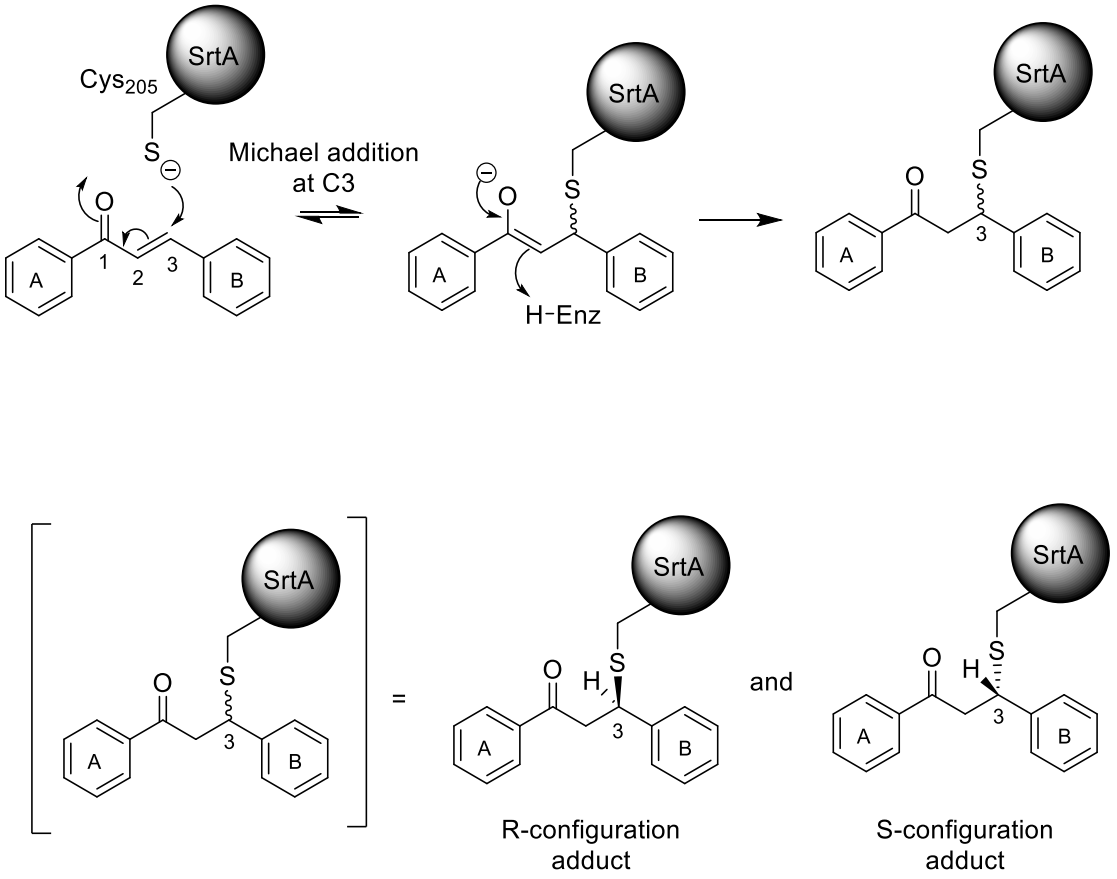


Figure 57: Proposed mechanisms for the reaction of *trans*-chalcone with SmSrtA. Michael adduct formed between trans-chalcone and SrtA, with nucleophilic attack of the cysteine residue across double bond of the α,β-unsaturated system to produce an adduct of mixed *R* and *S* configurations.

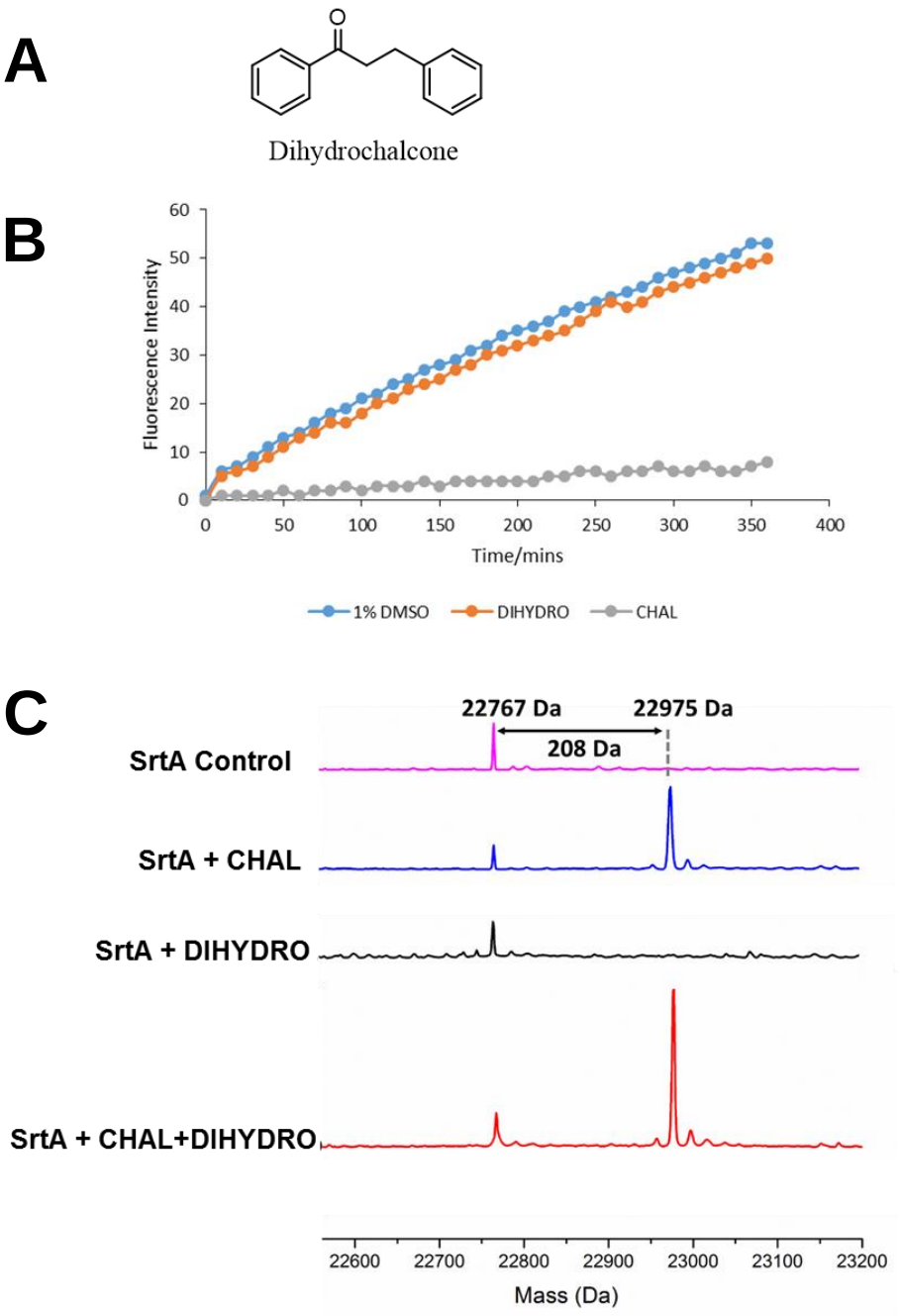


Figure 58: SrtA incubated with *trans*-chalcone and dihydrochalcone. (A) Chemical structure of dihydrochalcone. (B) Fluorescence intensity obtained from SrtA incubated with of *trans*-chalcone (100 µM) and dihydrochalcone (100 µM) for 16 h and analysed by FRET assay. (C) Mass spectrum of SrtA control (5 µM) incubated with (i) *trans*-chalcone (100 µM), (ii) dihydrochalcone (100 µM) and (iii) mixture of *trans*-chalcone (100 µM) and dihydrochalcone (100 µM) for 16 h.

4.3.5.2 Reversibility of the SrtA-modification

The MS analysis confirmed the formation of SrtA-chalcone adduct involving the active site residue, Cys205 via a Michael addition reaction. With extensive dialysis, enzyme activity was not restored, which suggests that adduct formation was irreversible. To further confirm that this reaction is irreversible, the inhibited enzyme was incubated with the reducing agent DTT (10 mM), which is expected to convert the modified Cys205 residue to its reduced state and restore activity, if the modification is reversible. In a previous experiment, it was observed that DTT also has the ability to increase the activity of SmSrtA by two-folds in the absence of an inhibitor molecule. Addition of DTT to SmSrtA-chalcone sample did not restore enzyme activity. However, the ~10% activity that remained in the inhibited enzyme was doubled in the presence of DTT, similar to that observed for the uninhibited enzyme (Figure 59A). MS analysis of the sample incubated with DTT showed a single peak of mass 22, 975 Da corresponding to the modified enzyme (SmSrtA-chalcone adduct) which further confirmed that the reaction of SmSrtA with *trans*-chalcone was irreversible (Figure 59B).

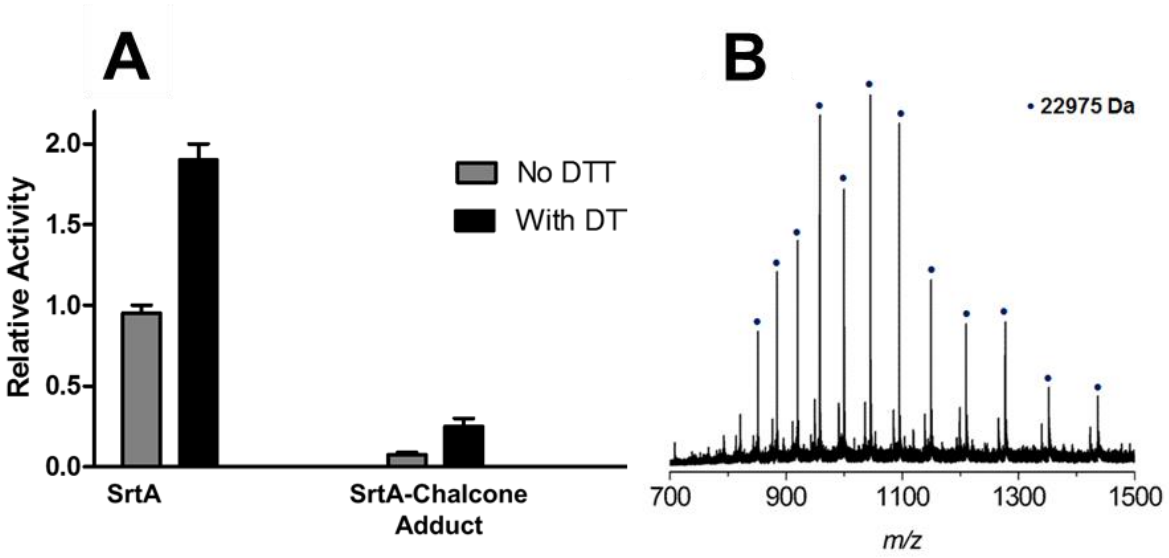


Figure 59: Effect of DTT on SmSrtA and SmSrtA-chalcone adduct. (A) SrtA (5 µM) was treated with *trans*-chalcone (100 µM) and incubated for 16 h. The overnight sample was incubated with DTT (10 mM) for a further 6 h, then dialysed against 20 mM MES, pH 6.5 and 125 mM NaCl and assayed for sortase activity. A SrtA control sample without the treatment of *trans*-chalcone was assayed with and without DTT (10 mM). (B) Positive ion mass spectrum of SrtA-chalcone adduct after incubation with DTT. A peak of mass 22, 975 Da was detected which corresponds to modified enzyme.

4.3.6 Structural Biology

All crystallographic analyses and molecular modeling in this chapter were carried out in collaboration with Dr. Jon Marles-Wright at the Institute of Structural and Molecular Biology at The Unversity of Edinburgh. The refined crystal structure and molecular models were also processed by Dr. Jon Marles-Wright.

4.3.6.1 Protein Crystallisation

Attempts were made to crystallise the purified WT enzyme, SmSrtA; however these crystal trials were unsuccessful due to the instability of the enzyme which observably underwent auto-proteolysis upon purification. However, the more stable SmSrtAH139A mutant was readily crystallised at a concentration of 11 mg/mL using the sitting drop

method. After 3 days, a single crystal was observed in the well solution containing 0.2 M $(NH_4)_2SO_4$, 25 % w/v PEG 4K and 0.1 M NaAc pH 4.6 (Figure 60A). Optimisation screens were prepared from these conditions by varying the percentage of each buffer component. The concentrated protein was crystallised by hanging drop vapour diffusion in drops of 2 µl protein plus 2 µl crystallisation solution (30 % w/v PEG 4000, NaAc pH 4.6, 0.2 M $(NH_4)_2SO_4$), over 1 ml of the latter within 5 days (Figure 60B). Several other conditions yielded crystals over 14-21 days (Figure 60 C-D).

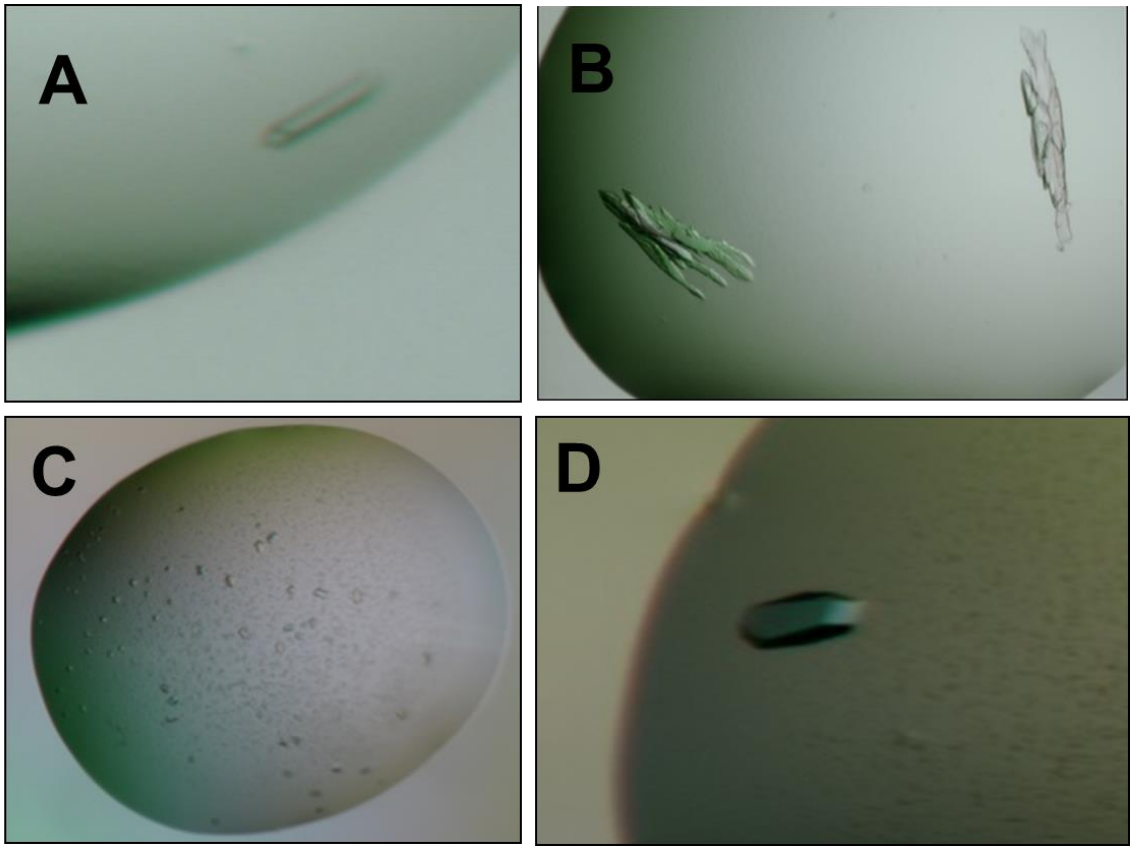


Figure 60: Crystals of SmSrtAH139A mutant. (A) Crystals were grown over 3 days in 0.2 M $(NH_4)_2SO_4$, 25 % w/v PEG 4K and 0.1 M NaAc pH 4.6. (B) Crystals grown over 5 days in 30 % w/v PEG 4000, NaAc pH 4.6, 0.2 M $(NH_4)_2SO_4$. (C) Crystals grown over 14 days in 0.1 M HEPES pH7 and 30 % Jeffamine Ed-2001. (D) Crystals were grown in the same condition as in C with twice the concentration of protein to buffer.

4.3.6.2 Data Collection and Structural Solution

The crystals of the SmSrtAH139A mutant diffracted to a resolution of 1.6 Å. They belong to the $P4_22_12$ space group with one molecule in the crystallographic asymmetric unit. The crystal structure of the enzyme (PDB code: 4TQX) was determined by molecular replacement using the structure of *S. pyogenes* sortase A, its closest homologue (66% homology), as the starting model (PDB code: 3FN5).[136] The refinement and statistics data for SmSrtAH139A are summarised in Table 18. The electron density map at residues 49-53 at the N-terminus was ambiguous; therefore the assignment of these residues remains undefined and they were omitted from the final model. This may suggest some degree of flexibility in the N-terminal domain. The final model of SmSrtAH139A contains 202 amino acids, inclusive of Ala41 to Glu48 and Lys54 to Phe246 (Figure 61). SmSrtA has a canonical eight stranded *β*-barrel core flanked by loops, coils and three α-helices.[183] The catalytic domain is located in the *β*-barrel core and is highly conserved among the other members of the sortase superfamily.[135] All of these structures have a similar relative orientation of the catalytic domain. The SmSrtA construct lacks the predicted forty amino acids of the trans-membrane domain, however, in addition to the conserved catalytic core (residues 88-206), the structure of SmSrtA also revealed a unique N-terminal α-helix (residues 69-89, relative to the full length enzyme). The helix is connected to the catalytic core by a short loop (residues 89-93) and is oriented in an extended position away from the catalytic domain.

Table 18: Data refinement and statistics for SmSrtA. Statistics for the highest-resolution shell are shown in parentheses.

Wavelength (Å)	0.92
Unit cell parameters (Å)	a=b=82.822 c=57.87, alpha=beta=gamma=90^{O}
Space group	P 42 21 2
Resolution range (Å)	47.44 – 1.37 (1.419 – 1.37)
I/σ (I)	25.19 (1.12)
Completeness (%)	99.85 (99.98)
Multiplicity	25.6 (26.0)
Total reflections	1094198 (109542)
Unique reflections	42782 (4216)
Wilson B-factor (Å^2)	19.25
R-merge	0.08007 (2.888)
R-meas	0.08172
CC1/2	1(0.618)
CC*	1(0.874)
Reflections used for R-free (%)	5
R-work	0.1450 (0.2806)
R-free	0.1792 (0.3296)
No. of non-hydrogen atoms	1837
macromolecules	1585
ligands	22
water	230
Protein residues	201
RMS (bonds) (Å)	0.009
RMS (angles) ^{O}	1.22
Ramachandran favored (%)	98
Ramachandran allowed (%)	2
Ramachandran outliers (%)	0
Clashscore	1.24
Average B-factor (Å^2)	27.50
Macromolecules (Å^2)	25.80
Ligands (Å^2)	49.60
Solvent (Å^2)	37.30

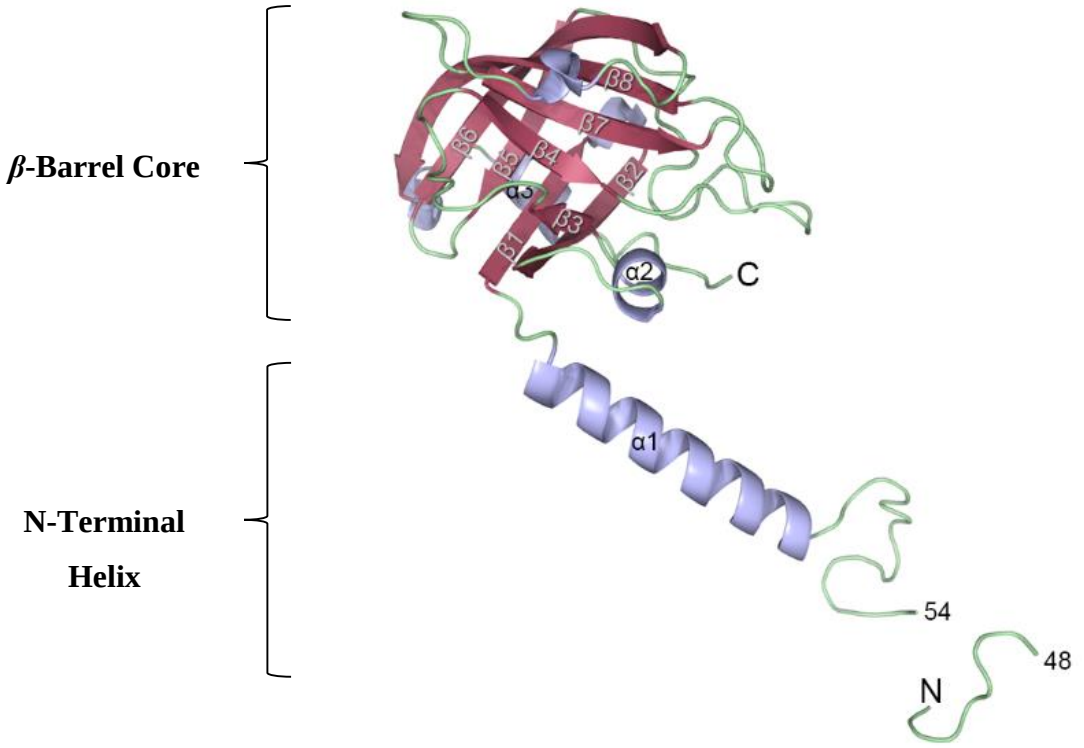


Figure 61: Crystal structure of SmSrtAH139A mutant (PDB code: 4TQX). (A) Cartoon model of the crystal structure of SrtA H139A mutant showing the catalytic domain made up of an eight-stranded *β*-barrel core and a unique *α*-helix at the N-terminus. Residues 49-53 remain unassigned. The structure was determined by molecular replacement with *S. pyogenes* SrtA (PDB: 3FN5).

This unique N-terminal helix has not been observed previously in other SrtA structures present in the PDB. An N-terminal helix is observed in the crystal structure of *S. aureus* SrtB (PDB code: 1NG5). However in contrast to the SmSrtA structure, this helix is hinged at Asp41 in SrtB and places it in an equivalent position to the C-terminus of SrtA and in close proximity to the *β*-barrel core (Figure 62A).[120] These structural differences may suggest a dynamic role of the N-terminal helix in positioning of the catalytic residues of sortases with its substrates for the efficient transpeptidation to take place. This study therefore reports the first structural evidence of a unique N-terminal domain of sortase enzymes which may function to position the catalytic core above the membrane where it can efficiently interact with its substrates.

The active site of SrtA enzymes consists of three highly conserved residues: His139, Cys205 and Arg214 which are essential for catalysis (numbering of residues is based on the sequence of SmSrtA). The crystal structure of the H139A mutant did not show any observable architectural compromise of the active site as a result of the mutation of His139 to an alanine residue (Figure 62B). The relative positions of the active site residues are also highly conserved. The only significant difference is seen between the orientation of the cysteine residue of *S. aureus* SrtB and the other sortase enzymes. In SaSrtB, Cys205 points towards the histidine residue in the active site, whereas in the other sortases the cysteine residue points away from His139 with a distance of ~7 Å.

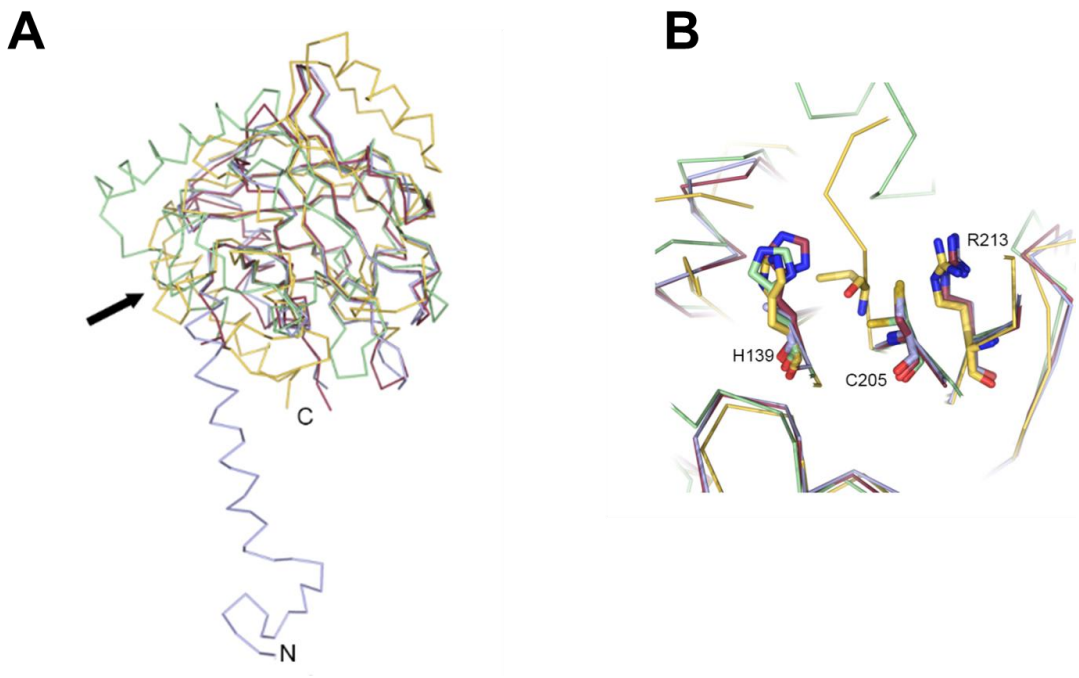


Figure 62: Overlay of sortases. (A) Stereo view of sortase structures showing the highly conserved core fold. (*Sm*SrtA, 4TQX, blue; *Sp*SrtA, 3FN5, raspberry; *Sa*SrtB, 1NG5, yellow; *Spn*SrtC, 2W1J, green). *Sm*SrtA has a unique N-terminal helix arrangement which is distinct to those seen for the other sortase family enzymes, N-terminus of *Sm*SrtA indicated with an 'N' and *Sa*SrtB indicated with an arrow. (B) Stereo view of the active site residues of the sortase structures as in (A), with residue numbers labelled for the *Sm*SrtA. The only significant difference in active residues is seen between the *Sa*SrtB active site cysteine and the other sortase family proteins, where it is shifted by approximately 4 Å due to a loop movement in the active site.

4.3.6.3 Molecular Modelling of SrtA with Symmetry-Element

In the crystal structure of the SmSrtAH139A mutant, Cys205, His/Ala139 and Arg214 are positioned on three neighbouring β-strands, β-7, β-4 and β-8 respectively which forms a tunnel-like hydrophobic catalytic pocket, similar to that described in other SrtA enzymes.[134] The catalytic pocket of SrtA is designed to accommodate the two bulky hydrophobic subtrates, the pentaglycine side chain of lipid II and a protein substrate consisting of the LPXTG motif at the N-terminal. Both substrates enter the active site from either side of the cleft to make contact with the catalytic triad (Cys-His-Arg). In addition to the natural peptide substrate, the hydrophobic cleft can also accommodate other molecules including peptides with similar surface properties. Interestingly, when viewing the unit cell of SmSrtA, the N-terminus α-helix (residues 69-89) from a symmetry-related molecule makes extensive contact with the active site cleft/binding groove of SmSrtA resulting in a "non-physiological dimer" (Figure 63). Two very significant contacts are observed between the aromatic side chains of Phe67 and Phe69 of the N-terminal helix and the active site. Both residues sit in close apposition to the active site cysteine and arginine residues, resulting in an obstruction of the active site cysteine residue (Figure 63 *inset*). These serendiptidous "non-physiological dimers" involving the active site offer a structural rationale as to why the WT SmSrtA undergoes auto-proteolysis. Although it does not contain a canonical "sortase motif" it appears that the N-terminus of SmSrtA binds to the active site and this self-recognition leads to amide bond cleavage.

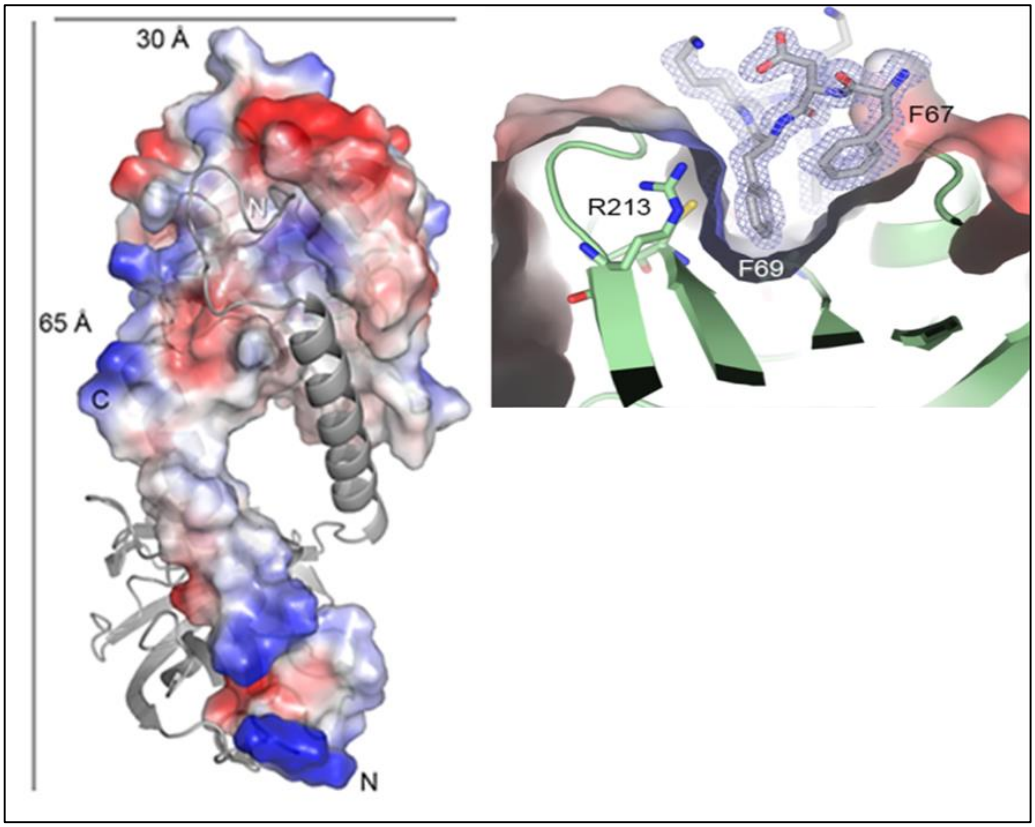


Figure 63: Modelling of *S. mutans* sortase A with symmetry-related molecule. Surface representation of the SmSrtA (H139A mutant) with symmetry related molecule. The N-terminal region of one molecule is bound to the other through extensive contacts and an extended helix occludes the active site cleft. (*Inset*) Experimental electron density map. Experimental 2mFo-DFc map contoured at 1.5 σ showing residues of the N-terminal helix within the active site cleft. Phe 67 and Phe69 sit in close apposition to the active site Cys and Arg residues.

4.3.6.4 Molecular Modelling of SrtA with trans-chalcone

Although milligram quantities of the chalcone-modified SmSrtA H139A enzyme were prepared, it was difficult to handle at high concentrations and failed to crystallise. Therefore a model of *trans*-chalcone bound in the active site of SrtA was built in order to understand the interactions within the binding cleft between the active site residues and the inhibitor. Model construction was guided by the the positions of the bound side chains of Phe67 and Phe69 from the symmetry-related SmSrtA partner, as well as the

crystal structures of the *S.aureus* SrtA-LPETG peptide complex (PDB: 1T2W) and *B.anthracis* SrtB-AEEK complex (PDB: 2OQZ). The 1T2W complex was chosen as it describes the substrate binding pocket where the scissile T-G bond of the peptide substrate would be positioned (Figure 64A).[134] Likewise the 2OQW complex describes the binding pocket of the AEEK inhibitor molecule that irreversibly forms a covalent adduct with the active site cysteine or SrtB (Figure 64B).[147]

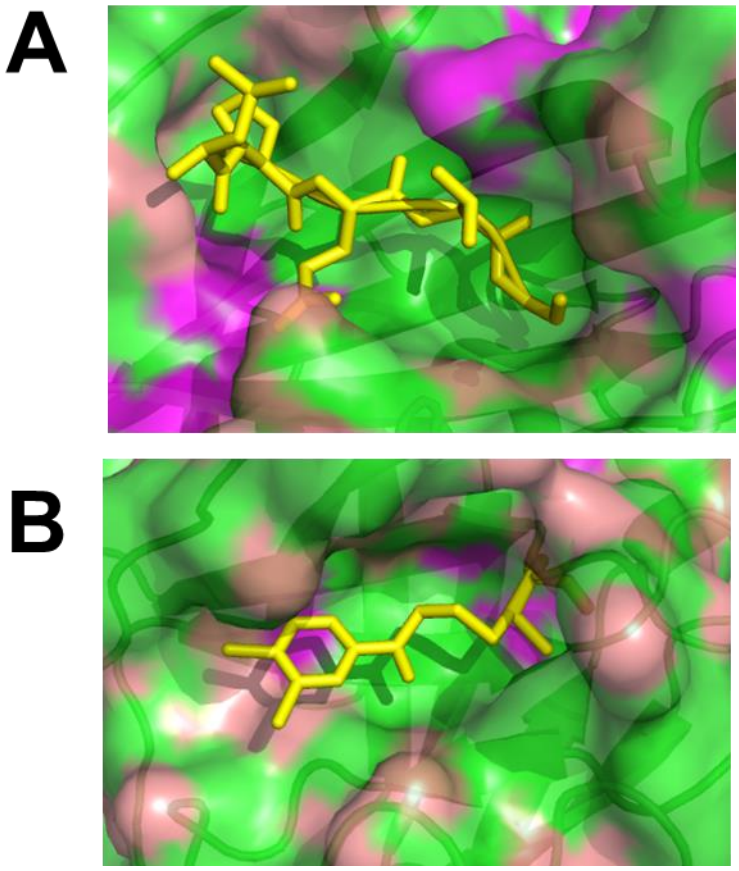


Figure 64: Surface plot representation of sortase showing binding cavity for substrates and inhibitors. (A) The LPETG peptide (coloured in yellow) buried in the deep binding pocket of SaSrtA. (B) The AEEK2 inhibitor (coloured in yellow) forms covalent adduct with active site cysteine residue and is buried in the peptide binding pocket of BaSrtB. The figure was generated using PyMol.

In the SrtA-chalcone adduct, a new bond is expected to be formed between the sulphur atom of the cysteine residue and the β-carbon of the chalcone molecule to give a

Michael product. The positions of the bound Phe67 and Phe69 phenyl ring side chains from the symmetry-related partner in the SmSrtA structure were used to place the two aromatic rings of the inhibitor covalently bound to the thiol of Cys205. The positions of rings A and B of chalcone in the active site can exist in two conformations to give both the *R* and *S* addition products (Figure 65). There appears to be no significant constrains on the orientation of *trans*-chalcone within the active site in the model, which makes free rotation possible about the C-S bond. In Figure 65, the red model shows the A-aryl ring of *trans*-chalcone bound in the LPXTG pocket and the B-aryl ring in the Lipid II pocket, while the blue model shows *trans*-chalcone with the B-aryl ring in the LPTXG pocket and the A-aryl ring in the lipid II pocket.

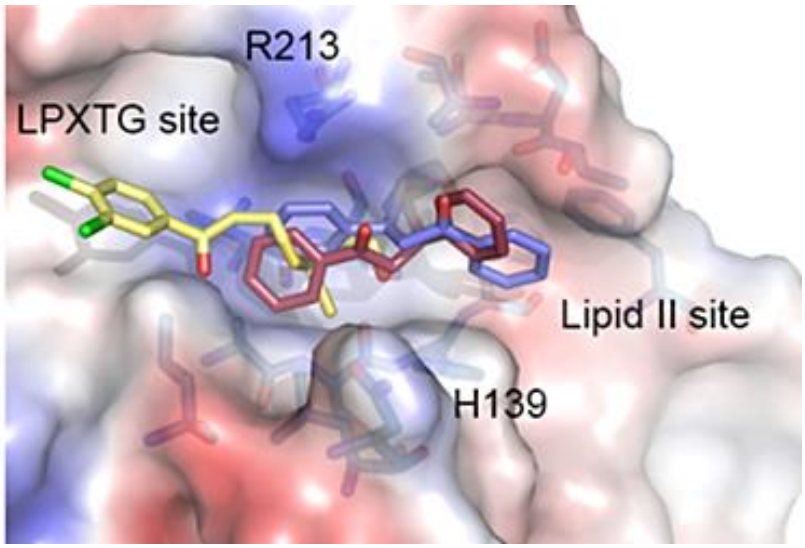


Figure 65: Modelling of SmSrtA with *trans*-chalcone. Model of the Michael adduct of *trans*-chalcone bound in the active site of SmSrtA in two different conformations. The red model shows the A-aryl ring of (*S*-form) *trans*-chalcone bound in the LPXTG pocket and the B-aryl ring in the Lipid II pocket. The blue model shows (*R*-form) *trans*-chalcone with the B-aryl ring in the LPTXG pocket and the A-aryl ring in the lipid II pocket.

A comparison of the SrtA-chalcone model with that of SrtB-AEEK structure (Figure 65, yellow model) reveals a difference in orientation of the active site residues. The active site cleft in SmSrtA is open at both ends leading to the peptide and the lipid II binding

sites, whereas in the SrtB a tyrosine residue, Tyr235 (which is absent in SrtA) obstructs one side of the binding cleft, after the SrtB-AEEK adduct is formed. These structural differences suggest a possible strategy for exploiting more structurally diverse chalcones from the flavonoid family as isoform-specific sortase inhibitors.

4.3.7 Antimicrobial Activity of *Trans*-chalcone

4.3.7.1 Biofilm Assay

SrtA modulates the cell-surface-related properties of *S. mutans* including the cell's biofilm formation ability.[128, 129] With strong evidence that *trans*-chalcone inhibits *S.mutans* SrtA *in vitro*, investigations were carried out to assess the effect of the inhibitor on the formation of biofilm by *S. mutans*. Using the AAA model, *S. mutans* biofilm was grown on a saliva-coated glass slide for 16-18 h in BHI medium supplemented with 0.2% sucrose in the presence of *trans*-chalcone (0-500 µM). The biofilm biomass was then stained with crystal violet and the optical density of the stain measured at 600 nm. A reduction in biofilm mass was suggestive of an inhibitory effect of *trans*-chalcone on biofilm formation. Morin, a known biofilm disruptor was used as a positive control. The data showed that biofilm formation was reduced in the presence of *trans*-chalcone in a concentration dependent manner up to 250 µM, with efficacy tailing off at higher concentrations (Figure 66). These data may suggest that there is a point of saturation of *trans*-chalcone with its biological target (s) in *S. mutans* biofilm, which may cause reduced growth of bacterial cells. At the same time, it could be that at high concentrations, *trans*-chalcone may induce a response which overrides the inhibitory effect of the molecule and causes bacterial cells to adapt a biofilm lifestyle for survival. Since biofilm formation is a complex and dynamic process that involves changes in microbial metabolism and signalling, this encouraging result suggests that the exact mechanism of chalcone inhibition *in vivo* requires further in-depth investigations to fully understand the observed trends in *S. mutans* biofilm growth.[184]

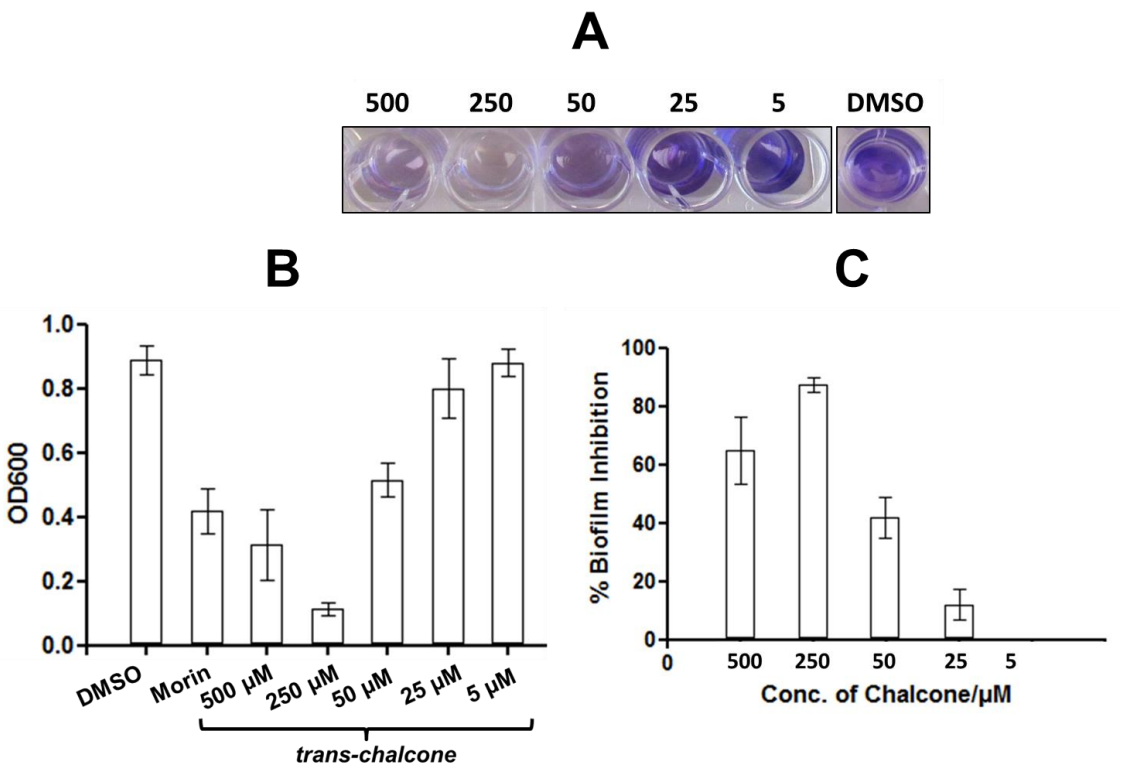


Figure 66: Effect of *trans*-chalcone on biofilm formation by *S. mutans*. (A) Photograph of crystal violet stain after destaining of *S. mutans* biofilms grown on saliva-coated glass slides in the presence of *trans*-chalcone. Cells were grown in BHI media containing *trans*-chalcone (0-500 µM). A stronger purple colour indicates that more biomass was present on glass slide before destaining as is seen in the wells containing 25 µM and 5 µM *trans*-chalcone. (B) Quantification of biofilm mass. The optical density of crystal violet stain was read at 600 nm. The values represent the average experiments performed in triplicate. DMSO was used as the negative control and morin (50 µM) as the positive control. (C) % Inhibition of biofilm formation in the presence of trans-chalcone. Biofilm formation was reduced in the presence of trans-chalcone in a concentration dependent manner up to 250 µM, with efficacy tailing off at higher concentrations. The OD_{600} for the DMSO sample was compared with the results for 250 µM *trans*-chalcone using the unpaired t-test. DMSO plot (n)=3, 250 µM trans-chalcone (n)=3, (df)=2, T value=24.1524, $p<0.0002$. This difference is considered to be extremely statistically significant.

4.4 Conclusion and Future Work

In this study, the structure of the H139A mutant of SrtA from *S. mutans* was determined which revealed a highly conserved catalytic core consisting of an eight *β*-stranded architecture. Additionally, the structure revealed a unique N-terminal domain consisting of an α-helix which may have an intrinsic flexible property necessary for the positioning of the active site with its substrates. It would be of interest to obtain the structure of the SrtA-chalcone adduct by setting up further crystal trials. Future work may involve the use of a thiol sepharose resin as a medium to immobilize unreacted SrtA after incubation with *trans*-chalcone, in an effort to obtain a higher concentration of pure modified enzyme. This structure would further confirm if the conclusions drawn from the SrtA-chalcone model are correct.

The inhibition studies carried out with *trans*-chalcone revealed the need to carry out further studies to screen other flavonoids for anti-sortase activity possessing a greater degree of specificity. This may also address the issue of identifying a molecule with high potency since *trans*-chalcone was only effective as an inhibitor at micromolar concentrations. *Trans*-chalcone was also found to possess antibiofilm properties against *S. mutans*, which suggests the potential application of the chalcone flavonoids as therapeutic agents for the control of biofilm-associated infections and highlights the need to carry out further studies to explore the mechanism by which biofilms are inhibited by this molecule. One possibility is that *trans*-chalcone, like other flavonoids may prevent the surface attachment of proteins involved in surface adherence of the bacterial cells via an inhition of the sortase enzyme responsible for this. This hypothesis can be further tested by performing western blot studies to identify these proteins in cell cultures grown in the absence and presence of *trans*-chalcone. In summary, the chalcone scaffold has provided insights into the mechanism of inhibition of SrtA by covalent modifiers and shows that the binding pocket of the enzyme can be further exploited for pharmacological benefits.

Chapter 5: Allicin-Containing Garlic Extracts

5.1 Introduction

Garlic extracts have been used for a long time for their antimicrobial properties. Several microorganisms have shown sensitivity to crushed garlic preparations. The fresh extracts consist of several thiosulfinates with allicin being identified as the major biologically active component accounting for up to 75% of total thiosulfinates.[96] Ajoene, a lipid-soluble allyl sulfide derivative of allicin, found in oil-derived garlic extracts has also been reported to have a broad spectrum of microbial growth inhibition.[185] Recently, ajoene has also been shown to inhibit quorum sensing in *P. aeruginosa*.[186, 187] Interestingly, we could find no published studies of the antimicrobial properties of garlic-related compounds on the *Burkholderia cepacia* complex (Bcc), a group of 17 closely-related species distributed widely in soil, water and the plant rhizosphere.[16] This is both surprising and ironic since as well as being important agents for bioremediation and biological control[188, 189] the Bcc are the major phytopathogens for allium species.[190]

In the last few decades, the Bcc has also emerged as an important opportunistic human pathogen, in particular as a cause of life-threatening lung infections in individuals with cystic fibrosis (CF) and chronic granulomatous disease.[17, 18] Although patient segregation and strict infection control have reduced the incidence of Bcc infections in individuals with CF, such infections remain an important clinical problem. At present, the most predominant Bcc species responsible for CF infections are *B. cenocepacia* and *B. multivorans*.[17, 22] Most transplant centres worldwide exclude individuals infected by *B. cenocepacia* from access to lung transplantation, the only proven treatment for severe CF lung disease. Thus, any new strategies that lead to the improved eradication of Bcc from an infected patient would be important. Unfortunately, a common feature of the Bcc is intrinsic resistance to most antibiotics[19]; hence antibiotic treatment presents a major challenge. At present, there is insufficient data to support the use of any specific antibiotic regimen against Bcc infection.[191, 192] There is an even greater need to have alternative therapies with the recent alarming report of a clinical trial of inhaled

aztreonam that failed to treat CF patients with *Burkholderia spp.* infection.[23, 24] Thus, the availability of novel antimicrobial agents against Bcc would represent a major clinical advance for patients with Bcc infection.

Bacterioferritin comigratory protein (BCP) belongs to a superfamily of peroxidase enzymes known as peroxiredoxins (Prxs). These enzymes carry out thiol-dependent reduction of peroxide substrates, which serve as a defense mechanism against reactive oxygen species (ROS). The BCP from *B. cenocepacia* (*Bc*BCP), was previously characterized by Clarke *et. al.* in a mass spectrometry study.[158] This enzymes protects *Burkholderia* from oxidative stress in a growth phase dependent manner via a 1-Cys catalytic pathway. This redox pathway was confirmed through experiments using iodoacetamide to covalently modify the free thiol groups resulting in the abolishment of the enzyme's peroxidase activity. The primary sequence of *Bc*BCP consists of only two cysteine residues, Cys-44 and Cys-98, which are highly conserved in the *Burkholderia* genome (Figure 67). Cys-44 was confirmed as the active-site peroxidatic residue in the thiol-dependent BCP by mass spectrometry, whereas Cys98 does not take part in catalytic activity. This cysteine residue (Cys-SH) is oxidized to a stable sulfinic acid (Cys-SOOH) on reaction with a peroxide substrate during the catalytic process. In the presence of a resolving partner, example, GSH, the oxidized enzyme is reduced to its native state.[158] Therefore any reactive thiol reagent, including allicin and other allium thiosulfinates are likely expected to react with the free thiols of BCP and also abolish its peroxidatic activity.

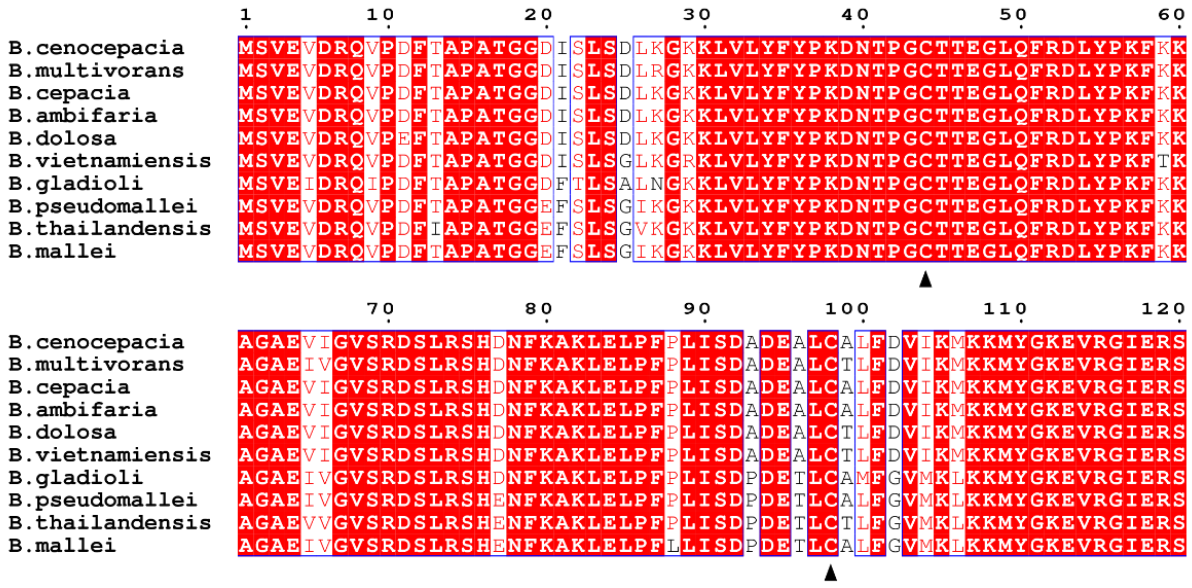


Figure 67: Sequence alignment of BCP from *Burkholderia*. Sequences in red show universally conserved residues in *Burkholderia* BCP. The black triangle highlights the two highly conserved cysteine residues of BCP.

5.1.1 Aims and Objectives

Given the dearth of new antibiotics from conventional approaches within the pharmaceutical industry, the development of novel antimicrobial compounds derived from natural sources would be welcomed [29, 30]; this is particularly relevant to the management of lung infections which are the primary cause of morbidity and mortality in the CF population. As previously indicated, the therapeutic potential and complex chemistry of alliums is well documented but remains enigmatic.[159, 193] Therefore the antimicrobial potential of allicin-containing plant extracts against Bcc is worth investigating.

Before this study, the antimicrobial activity of freshly prepared aqueous garlic extract (AGE) and aqueous allicin standard (AAS) against Bcc was unknown. Therefore this study was designed to investigate the antimicrobial potential of AGE and AAS against Bcc by employing a multidisciplinary approach using both analytical chemistry and

microbiology. This required an accurate and consistent quantification of the amount of allicin present in the freshly prepared garlic extracts in relation to aqueous allicn standards by HPLC. From this analytical analysis of AGE and AAS, determination of their MICs and MBCs would give some indication of their antimicrobial activities. Another objective of this study was to determine the efficacy of various assays including broth/agar dilution, paper disc and agar well diffusion in analyzing the antimicrobial potential of AGE and AAS against Bcc.

The exact chemical mechanism(s) by which allicin exerts its antibacterial activity has been the subject of numerous studies, which suggest that allicin reacts with free thiols in important bacterial components essential for life. Indeed, allicin is known to covalently modify the free amino acid cysteine via the formation of allyl-disulfide species.[194] Given previous evidence that allicin preferentially interacts with cysteine, it has been hypothesized that the allylthio moiety of allicin allows it to react with cysteine-containing *Burkholderia* enzymes involved in key biosynthetic pathways. The exact targets of allicin are not known but it can be assumed that there will be many proteins that have the potential to be modified. As a relevant model protein, the Bcc-derived peroxidase enzyme, BCP, which has been previously shown to play a role in the detoxification of reactive oxygen species (ROS) within *B. cenocepacia* was used to study the reaction of allicin with thiol-containing proteins within the *B. cenocepacia* genome through time dependence and reversibility studies, using high resolution mass spectrometry.

5.1 Results and Discussion

5.2.1 Analysis of Garlic and Allicin Samples

5.2.1.1 Allicin Standard

An aqueous solution of pure allicin (AAS) obtained commercially was prepared and used as the primary standard for determining the allicin content in fresh garlic extract. The purity of AAS was confirmed using a published HPLC method with UV detection at 240 nm on a C18 reverse phase column.[160] Allicin eluted with a retention time of 4.1 min and an observed *m/z* of 185.0067 (consistent with the $[M+Na]^+$ species; theoretical *m/z* 185.0065; $[C_6H_{10}OS_2+Na]^+$). Only the peak for allicin was detected with no appreciable degradation component based on the spectrum (Figure 68A). Allicin standards were prepared by dissolving allicin in sterile, distilled water to produce aqueous solution, ranging from 8-1000 µg/mL by serial dilution. An injection volume of 20 µL of the standards resulted in a range of 0-20 µg of allicin being loaded onto the column. A standard curve for these solutions gave a straight line with a correlation coefficient of r^2 = 0.9987 and gradient 0.57 ± 0.01 mAU.min. μg^{-1} (Figure 68B).

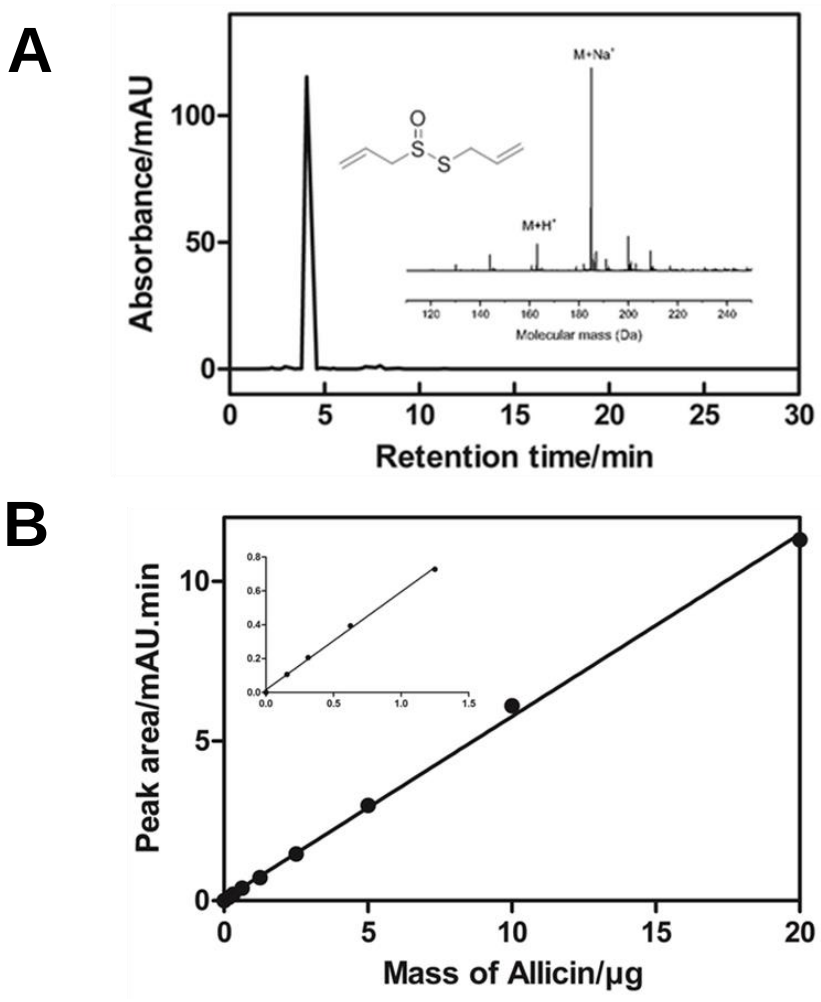


Figure 68: Analysis of standard aqueous solutions of allicin. (A) Chromatogram of AAS (1mg/mL) as obtained by HPLC with UV detection at 240 nm. Allicin elutes at 4.1 mins from a C18 reverse phase column at a flow rate of 1 mL/min using a linear gradient of MeOH:water (60:40). (B) Calibration curve of AAS as obtained by HPLC-UV analysis. Standards ranged from 8-1000 µg/mL. Each run was done in triplicate. A correlation coefficient of r^2 = 0.9987 and gradient 0.57 ± 0.01 mAU.min. μg^{-1} were obtained.

5.2.1.2 Aqueous Garlic Extracts

Aqueous garlic extract was prepared from homogenizing garlic bulb in sterile, distilled water (1 g of garlic bulb per 1 mL of water). The freshly prepared garlic extracts were also analysed using the same HPLC conditions described above for AAS. The elution profile of AGE revealed a complex mixture with several well-resolved peaks including a sharp peak for allicin at 4.1 mins (Figure 69). The thiosulfinates present in aqueous garlic homogenates have been previously characterized and quantified in a study conducted by Lawson *et al.*[96] The amount of allicin in freshly prepared AGE, based on the standard curve, was determined as 67.6 ± 7.9 µg in 20 µl, which equates to 3.38 mg ± 0.39 mg/mL (allicin/AGE, w/v). This concentration is comparable to the amount reported by Fugisawa *et al*. (1.32 ± 0.43 mg/mL) and other literature reports of the amount of allicin in different strains of garlic which ranges from 2.8-7.7 mg per g of garlic.

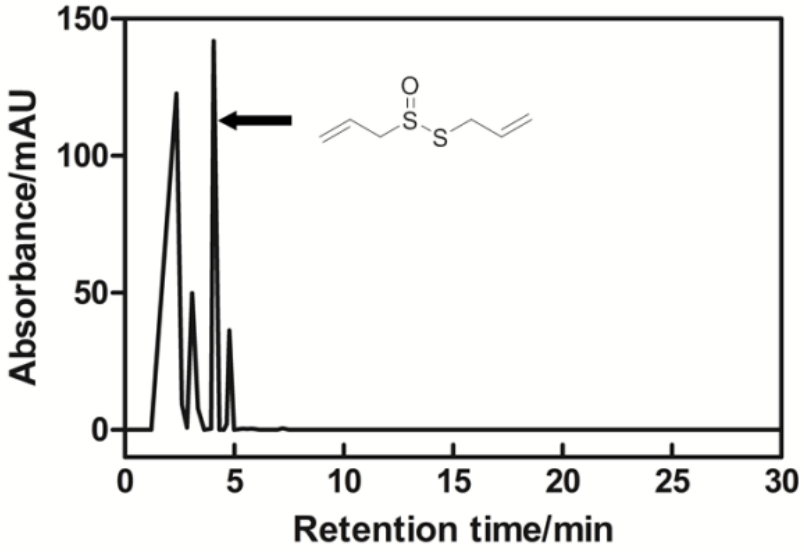


Figure 69: Analysis of AGE by HPLC. The chromatogram of neat AGE (1 g of garlic bulb per 1 mL of water) obtained by HPLC with UV detection at 240 nm shows a complex mixture. Allicin elutes at 4.1 mins from a C18 reverse phase column at a flow rate of 1 mL/min using a linear gradient of MeOH:water (60:40).

5.2.1.3 AllicinMax™

A number of garlic supplements e.g. AllicinMax™ are on the market with claims that they contain similar amounts of active compounds as fresh garlic extracts in particular allicin, the main biologically active component of garlic. Since allicin does not exist naturally in garlic, these supplements normally contain allinase and the precursor molecule alliin which is enzymatically converted to allicin after consumption. However evidence from a study carried out by Lawson showed that many commercially available garlic supplements show poor release of allicin which may be partly due to the enzymatic conversion being significantly affected by gastric acid, intestinal proteases as well as the manufacturing process.[160, 195, 196] With the aim of identifying pharmaceutical formulations of garlic extracts with potent antimicrobial activities similar to that of pure allicin and fresh garlic extract, an investigation of the bactericidal activity of the garlic supplement AllicinMax™ against Bcc was carried out. This particular supplement was chosen as the manufacturer's claim suggests that each capsule contains 180 mg of 100% stabilised allicin which is converted to allicin derivatives upon consumption. Extraction of the allicin was attempted by dissolving the capsules in 0.1 M HCl at 37 ^{o}C for 1 hour followed by the neutralisation of the extract with 0.1 M NaOH. The extract was analysed by HPLC with UV detection using the same protocol outlined for the analysis of AAS and AGE. HPLC and mass spectrometry analysis of AllicinMax™ revealed a complex and diverse composition of chemical compounds with no quantifiable amounts of allicin detected (Figure 70).

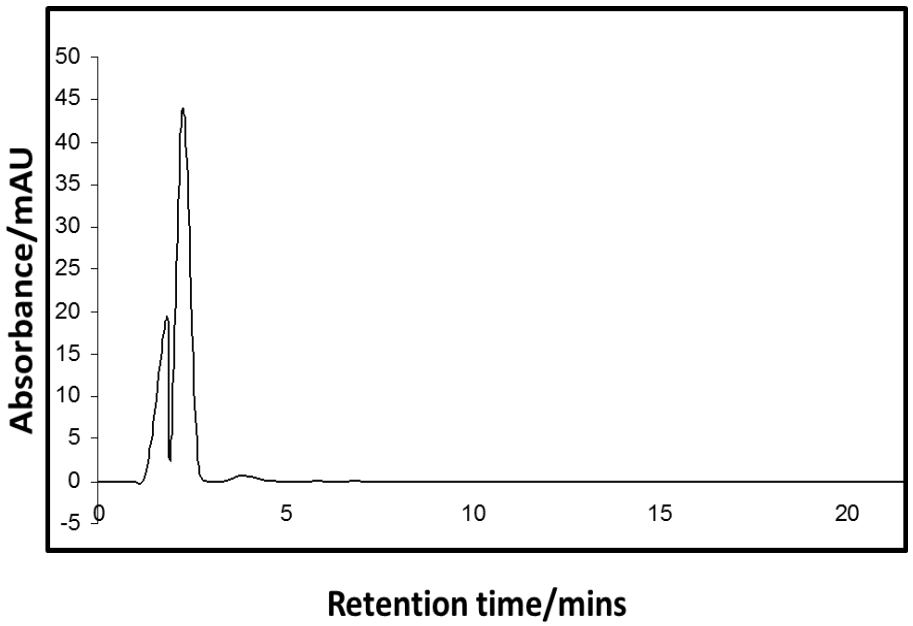


Figure 70: Chromatogram of AllicinMax™ extract analysed by HPLC with UV detection at 240 nm. Extract was prepared from the content of 4 capsules. No quantifiable amounts of allicin detected at the retention time of 4.1 mins.

5.2.2 Antimicrobial Activity

5.2.2.1 MIC for AGE

The antimicrobial activity of freshly prepared AGE was assessed against 38 Bcc isolates, representing nine Bcc species and which included 30 isolates from an experimental Bcc strain panel. [197] The MICs were determined by the agar dilution method and gave results ranging from 0.5 to 3% (Table 19), where 100% equates to 10g of garlic bulbs homogenised in 10 ml of sterile, distilled water. Interestingly, comparison of the antimicrobial activity of independent AGE preparations showed little batch-to-batch variation, which correlates well with the high batch to batch reproducibility of the measured allicin content of AGE. Activity also remained stable when AGE was stored at 4 ^{o}C for seven days or for at least 6 months at -80 ^{o}C. Aqueous solutions of allicin have been reported by Lawson to be very stable for up to two years at -80 ^{o}C,[195] which could

be due to the hydrogen bonding between water and the reactive oxygen atom of allicin.[107] This initial survey of the antimicrobial activity of AGE against Bcc confirmed sensitivity of the microbes to the extract. It was noted that the epidemic, clinically-relevant isolate *B. cenocepacia* C3168 was very sensitive to AGE (MIC = 0.5%). The amount of allicin present in garlic preparations used in this study can be quantified from these results. Using the MIC of 0.5 % for AGE against the *B. cenocepacia* C3168 (Edinburgh Lab No., C6433), this dilution of AGE equates to ~16.9 µg/mL of allicin.

Table 19: MICs (%) of AGE for Bcc. MICs were determined by the agar dilution method and gave results ranging from 0.5 to 3%, where 100% equates to 10g of garlic bulb homogenised in 10 ml of sterile, distilled water.

Strain	Edinburgh Lab No.	LMG No.	Strain Name	AGE (%)
B.cepacia	J673	1222	ATCC 25416	2
B.cepacia	J675	2161	ATCC 2161	1
B.cepacia	C2970	17997		0.5
B.cepacia	C3159	18821	CEP509	0.5
B.multivorans	C1576	16660		0.5
B.multivorans	C1962	16665		0.5
B.multivorans	C3160	18822	C5393	1
B.multivorans	C3161	13010		1
B.multivorans	C3162	18825	CF-A1-1	1
B.multivorans	C3163	18824	JTC	0.5
B.multivorans	C3164	18823	249-2	0.5
B.multivorans	J2511	17588	ATCC 17616	1
B. cenocepacia	J415	16654		1
B. cenocepacia	C1394	16659		1
B. cenocepacia	J2315	16656		0.5
B. cenocepacia	J2524	18832	ATCC 17765	3
B. cenocepacia	C3165	18826	BC7	0.5
B. cenocepacia	C3166	18863	K56-2	1
B. cenocepacia	C3167	18827	C5424	0.5
B. cenocepacia	C3168	18828	C6433	0.5
B. cenocepacia	C3169	18829	PC184	0.5
B. cenocepacia	C3170	18830	CEP511	1
B.stabilis	C3171	14294		0.5
B.stabilis	C3172	18870	C7322	0.5
B.stabilis	C3173	18888		1
B.stabilis	C3174	14086		1
B. vietnamiensis	C2978	16232		1
B. vietnamiensis	C3175	18835	PC259	1
B. vietnamiensis	C3176	18836	FC441	3
B. vietnamiensis	C3177	10929		1
B.dolosa	J3358			1
B.dolosa	J3361			1
B.ambifaria	J3362			2
B.ambifaria	J3362			1
B.anthina	J3364			1
B.anthina	J3366			1
B.pyrrocnia	J3368			2
B.pyrrocnia	J3369			0.5

5.2.2.2 MICs and MBCs for Allicin

Since Bcc showed sensitivity to AGE, the next objective was to investigate the antimicrobial activity of AAS. In light of the high cost of pure allicin (€150 per mg), the Bcc panel for screening was restricted to five isolates including three representatives of *B. cenocepacia*, the most clinically relevant Bcc strain. In addition to MIC determination, MBC assays for AAS were also performed on this smaller panel using the microtitre broth technique recommended by the NCCLS.[171] The MICs for AAS ranged from 8-62 µg/ml, with two *B. cenocepacia* strains, J2315 and C6433, being the most sensitive of the panel. The MBCs for AAS ranged from 31-62 µg/ml (Table 20).

Table 20: MICs (µg/mL) and MBCs (µg/mL) of AAS for Bcc. The Bcc panel for screening was restricted to five isolates including three representatives of *B. cenocepacia*.

Strain	Edinburgh Lab No.	LMG No.	MIC	MBC
B. cenocepacia	J2315	16656	8	31
B. cenocepacia	C6433	18828	8	31
B. cenocepacia	C3165	18826	16	31
B. stabilis	C3172	18870	16	31
B.cepacia	J673	1222	62	62

This analysis allows for the comparison between the MIC of pure allicin (AAS) and AGE. For instance, the MIC of pure AAS against the *B. cenocepacia* isolate C6433 was determined as 8 µg/mL allicin and the MIC of AGE 0.5%, which equates to ~16.9 µg/mL of allicin in the garlic extract. These values suggest that, against this isolate, allicin is the major antibacterial species in the AGE preparation. This study therefore reports the first evidence for the antimicrobial activity of allicin against the Bcc.

5.2.2.3 Plate Diffusion Assay

The antimicrobial activity of AGE was confirmed using impregnated paper discs and agar well diffusion assays. This experiment was carried out to provide further information on other suitable screening assays. An inoculum of 10 µl (10^6 CFU/ml) was used per disc or well. Antimicrobial activity was determined by a zone of inhibition after overnight incubation of the microbes with AGE or AAS at 37 ^{O}C. Both assays showed that *B.cenocepacia* was sensitive to AGE, confirming what was previously observed using the agar dilution method. No mutant strain resistant to allicin or AGE was observed in the inhibition zones against *B. cenocepacia* C3168 on the agar plate (Figure 71).

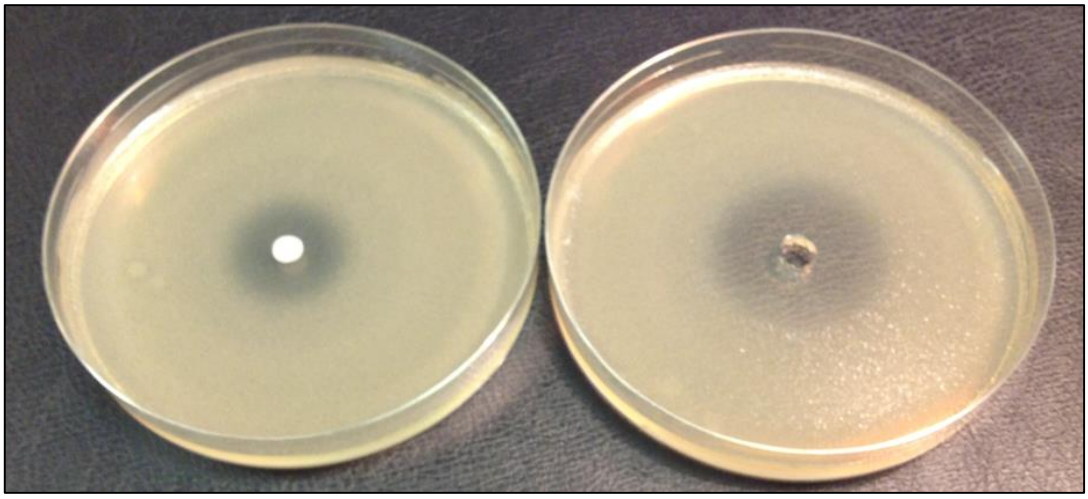


Figure 71: Zones of inhibition. Antimicrobial activity against *B. cenocepacia* C3168 demonstrated on seeded agar plates by AGE is indicated by clear zones where bacterial growth is inhibited. Freshly grown *B. cenocepacia* C3168 (10^6 CFU/mL) were embedded in 20 mL of warm agar in a Petri dish. Afterwards, a hole was punched out, 0.6 cm in diameter (right) and filled with AGE or a paper disc impregnated with AGE was placed in the centre of the plate (left) and incubated at 37 °C 16-18 h.

5.2.3 Interactions of AAS and AGE with *Burkholderia* BCP

5.2.3.1 Expression and Purification of BCP

The recombinant WT BCP from *B. cenocepacia* J2315 (referred to here as BCP) was expressed in an *E. coli* BL21 (DE3) host with an N-terminal hexa-histidine tag as described by Clarke *et. al*.[158] The protein was purified to homogeneity by nickel affinity chromatography to give pure sample that could be used directly for analysis without further purification by SEC. A small fraction of the sample was subjected to oligomeric analysis using an analytical grade Superdex 200 10/300 GL column, which gave a symmetric peak representing the monomeric enzyme with a mass of ~19 kDa. The mass of the protein was further confirmed by SDS/PAGE analysis and ESI mass spectrometry, resulting in an average mass of 18992 Da (Figure 72).

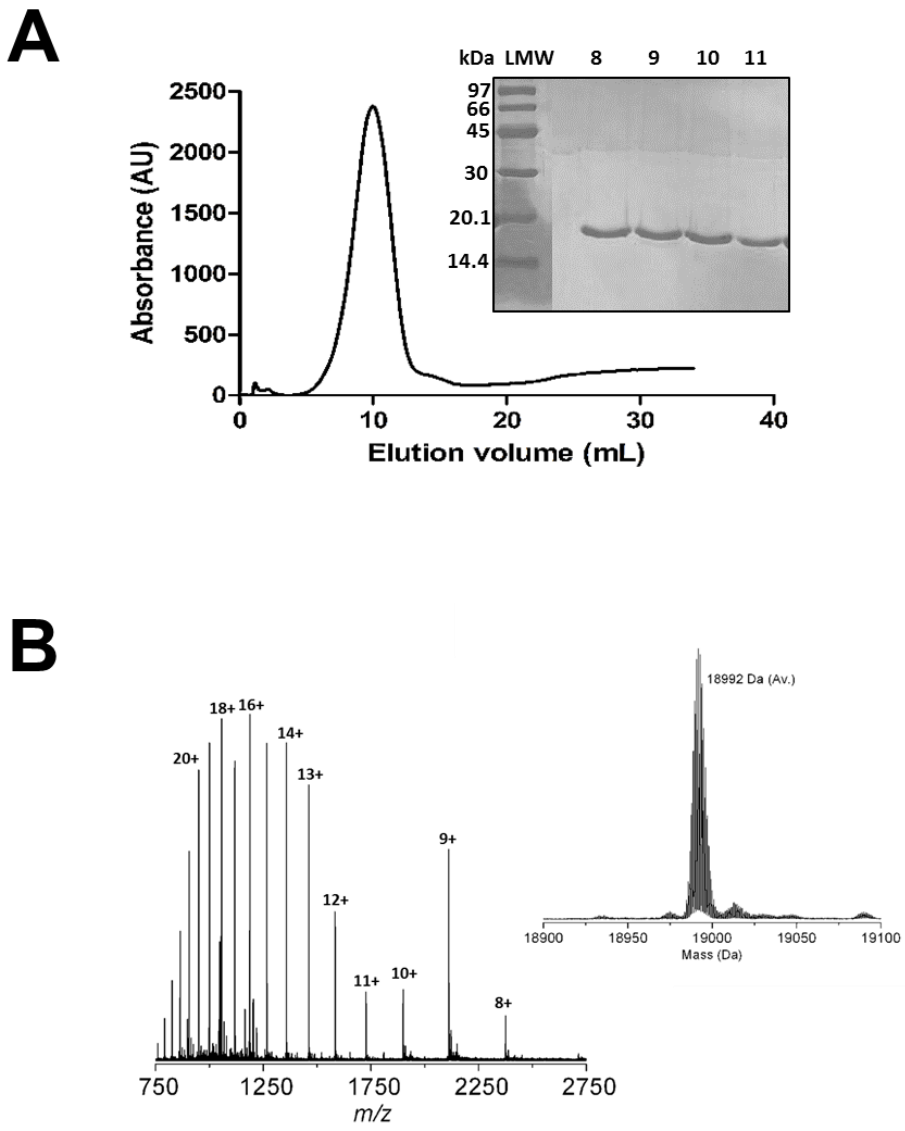


Figure 72: Purification and MS analysis of *B. cenocepacia* BCP. (A) Elution profile of BCP from analysed by SEC to give a monomeric enzyme. SDS/PAGE analysis (*inset*) shows the mass of the enzyme to be ~19 kDa; lane 1: low molecular weight marker; lane 2-4: fractions 8 to 10 for the peak that eluted in ~10 mL of elution buffer (50 mM Tris-HCl, 500 mM NaCl, pH 7.5) (B) Charge state distribution obtained for BCP. This corresponds to a deconvoluted molecular mass of 18992 Da (*insert*), which was consistent with the predicted theoretical mass based on the amino acid sequence of the protein (theoretical mass: 18992 Da).

5.2.3.2 Reactions of BCP with AAS and AGE

B. cenocepacia BCP has a mass of 18992 Da which confirms that both cysteine residues are in the reduced form (Cys-SH). N-gluconoylation is very commonly observed in proteins with an N-terminal histidine tag expressed in *E. coli*, this phenomenon was observed in purified BCP, evident by a small percentage of the purified enzyme having an additional mass of 178 Da (Figure 73A).[198] Incubation of 20 µM BCP with 1 mM allicin led to the formation of mixed allyl-disulfide species with masses of 19064 and 19136 Da after 60 mins. These observed increases in mass (+Δ72 and +Δ144 Da) are consistent with the addition of a single allyl thiol group from allicin (C_3H_5S) to either one or both cysteine residues (Figure 73B). Incubation of modified BCP with the reducing agent TCEP at 1 mM resulted in reduction of the enzyme to its native unmodified state, which suggests the reaction between BCP and allicin is reversible. This result is not surprising as allicin is expected to react with the free thiols on BCP in a non-specific manner as shown by previous studies with cysteine, glutathione and papain.[194] Incubation of BCP with AGE for 60 mins followed by repurification of the enzyme by nickel affinity chromatography also resulted in a similar mass increase as seen with pure allicin, which suggests that AGE and the AAS react with the protein target in the same manner (Figure 73C). The results from the proteomic analysis of BCP with the AGE were unexpected, as garlic extract is a complex mixture as observed from our HPLC analysis. This clearly-resolved data confirms that allicin is the main component of freshly prepared garlic that reacted with BCP. More importantly, this is the first study where AAS and AGE have been analysed by protein mass spectrometry with a relevant protein target from a bacterial pathogen. This study opens the door to apply this powerful analytical tool for identifying many of the protein targets that are modified by allicin within the proteome of *Burkholderia* and other pathogens. It could also prove useful to study other similar allylthio-containing natural products like ajoene derived from garlic with relevant bacterial protein targets.

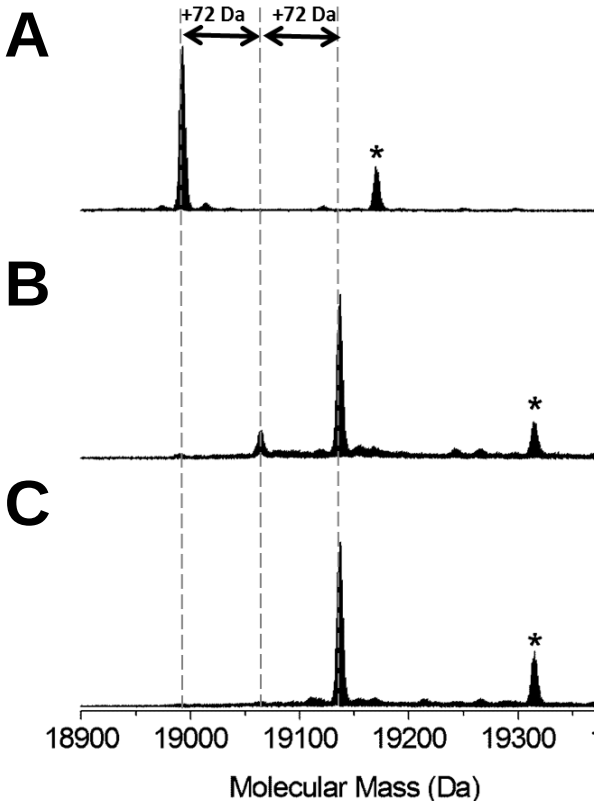


Figure 73: MS analysis of *Burkholderia* BCP incubated with AAS and AGE. (A) Purified *Burkholderia* BCP has a mass of 18, 992 Da, consistent with its predicted mass based on the amino acid sequence. (B) *Burkholderia* BCP after incubation with 1 mM AAS for 1 h results in the formation of two covalently modified species of mass increases +72 Da and +144 Da. (C) *Burkholderia* BCP after incubation with AGE for 1 h results in the formation of a covalently modified species of mass increase +144 Da corresponding to the addition of two allicin adducts. * *denotes enzyme species that undergo α-N-gluconoylation during protein expression.*

5.2.3.3 Reaction Mechanism

Allicin and AGE react with both Cys44 and Cys98 residues of BCP to give modified isoforms of the enzyme with masses 19, 064 Da and 19, 136 Da. These observed mass increases of 72 and 144 Da correspond to the addition of one and two allylthio groups respectively by reaction with allicin. The proposed mechanism for the formation of the

BCP-allythio derivatives involve nucleophilic attack of a BCP cysteine thiol (either Cys44 or Cys98) on the more electrophilic sulfur of allicin to produce a new disulfide bond. A second molecule of allicin undergoes a similar reaction to add a second allylthio group, producing a sulfenic acid byproduct (Figure 74). Reduction of the modified BCP with excess reducing agent (tris(2-carboxyethyl) phosphine, TCEP) returned the enzyme to the unmodified form (18992 Da) which suggests that the reaction of allicin with BCP is chemically reversible in the presence of strong reducing agents.

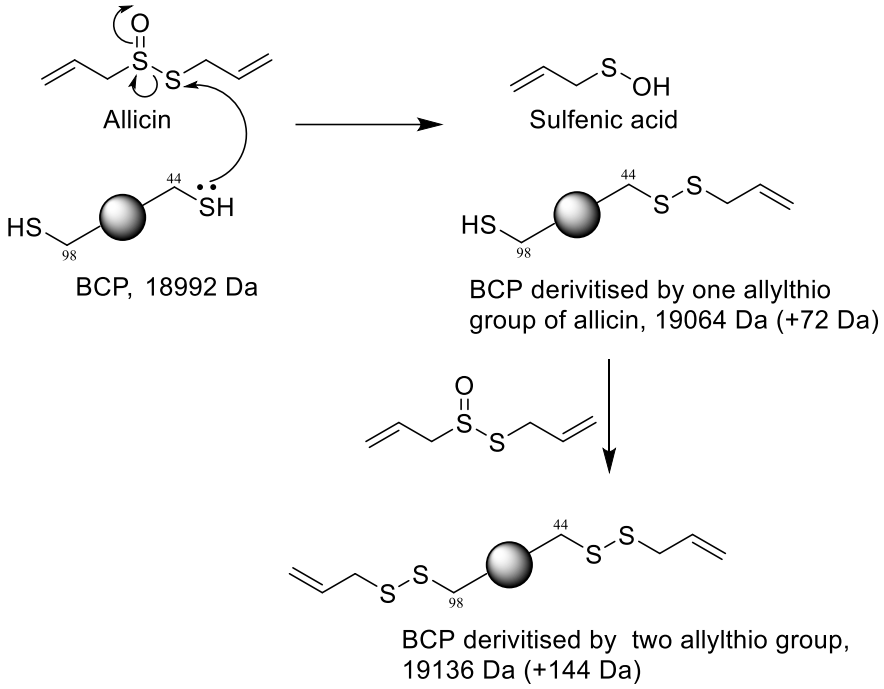


Figure 74: Proposed mechanism for the reaction of allicin with BCP. Formation of the BCP-allythio derivative involves nucleophilic attack of a BCP cysteine thiol (either Cys44 or Cys98) on the more electrophilic sulfur of allicin to produce a new disulfide bond.

5.3 Conclusion

In this study the allicin content of freshly prepared AGE was determined by HPLC analysis using pure allicin standard for quantification. The data from this analysis were consistent and correlated well with the trends observed in the antimicrobial activities of allicin and AGE. Together these results confirmed that the primary bacterial killing activity of fresh garlic extracts against *B. cenocepacia* lie with allicin. Production of antimicrobial compounds by plant alliums which target *Burkholderia* species, their principal bacterial pathogen, is not surprising. However, what is surprising is that, to our knowledge; such activity has not been reported or investigated previously. Therefore this study reports the first evidence for the antimicrobial activity of allicin against the Bcc, with killing activity against *B. cenocepacia*, the most prevalent and transmissible Bcc species isolated from CF infections. This finding is very significant as Bcc isolates exhibit intrinsic resistance to antimicrobial peptides and other antibiotics.[19, 199, 200] This novel investigation suggests that further work is required to comprehensively describe the antimicrobial mechanisms of allicin and to assess allicin-containing formulations as adjuncts to conventional antimicrobial agents presently used against this challenging group of bacterial pathogens.

Previous investigations of novel antimicrobial compounds against Bcc species have included mushroom extracts[201], polyketides[26], docosahexaenoic omega- 3 fatty acid[202] and microbe-derived volatile organic compounds[28]. Our results demonstrating the inhibitory activity of allicin-containing garlic extracts against *Burkholderia* species provide another potential line of research that is particularly relevant to Bcc infection. Evidence from animal and human studies have shown that allicin-containing garlic homogenates, delivered orally, intravenously or intraperitoneally, act as potent antioxidants leading to reduction of neutrophil-mediated lung damage, a major mechanism of Bcc lung infection.[203] The stability of the antimicrobial properties of AGE during storage at 4°C and −80°C coupled with the batch-to-batch reproducibility were notable. Similarly, the rapid therapeutic action of garlic homogenates in serum and

plasma in the treatment of cardiovascular disorders[204, 205] could be a particular advantage in the treatment of the life-threatening septicaemia that is characteristic of cepacia syndrome. Our investigation suggests that further work is required to comprehensively describe the antimicrobial mechanisms of allicin and to assess allicin-containing formulations as adjuncts to conventional antimicrobial agents presently used against this challenging group of bacterial pathogens.

Chapter 6: General Conclusion

With the impact of infectious diseases on the rise, there is a great demand for new antimicrobial therapies through the identification of lead compounds and novel antimicrobial targets. Current research has shown that NPs are still relevant sources for identifying lead compounds due to their complex functionalities and diverse chemical structures. Additionally, there are numerous gene clusters from natural sources, for which the corresponding NP has yet to be identified.[1, 206] The connection between genes, enzymes and molecules are now being exploited through the current advancements in fast and inexpensive genome sequencing technologies, which are increasing our knowledge of the biochemistry involved in biosynthetic pathways of many plants, fungi and bacteria, the main sources of NPs. From the study of biosynthetic pathways, new classes of enzymes are emerging rapidly, with novel structures and functionalities, which have provided insight into the source of the diversity of NPs. The varied functionalities within the AOS family of enzymes are a typical example, as several natural products essential for and detrimental to life have emerged from their activities. In this thesis, we reported the characterization of the putative AOS, annotated as TamD from the tambjamine biosynthetic pathway in *P. tunicata*.

Tambjamines are known to be potent biologically active molecules, with antimicrobial and anticancer properties. However, not much is known about the enzymes involved in the biosynthesis of these molecules. The TamD enzyme plays a very important role, existing as a didomain protein, incorporating both a malonyl unit and a seryl unit into the pathway for the formation of a bipyrrolic intermediate. The source of the enzyme's specificity for L-serine remains unknown, which serves as a point of interest for future bioengineering to introduce additional diversity within the family of tambjamines, since the AOS enzymes are known to all utilize specific amino acids in their PLP-dependent condensation reactions. Additionally, the structure of TamD is very fascinating in that it bears an unusual ACP domain at the N-terminus that is absent in the other well-characterised AOS enzymes (KBL, SPT, AONS and CqsA). It was of interest for us in

this study to investigate the topological relationship between the ACP and the AOS domains of TamD, which would provide insight into how the other AOS enzymes interact with their respective ACP partner acyl substrates. This information is very important when it comes to further engineering of the enzyme. Due to flexible and dynamic nature of the N-terminal domain, our efforts to obtain a high resolution crystal structure of TamD were hindered. Future efforts to capture this important structure would therefore involve the production of a substrate analogue that can link both the ACP and the AOS domains in order to reduce disorder of the flexible residues within the enzyme and increase the likelihood of crystallization and diffraction. Also solving the ACP and AOS domains independently may result in further information of the topological relationship of both domains to each other.

NPs are not only useful in identifying new leads and studying biosynthetic pathways. Existing NPs with known biological activities can also serve as pharmacological probes to identify novel antimicrobial protein targets and assess their underlying mechanisms of inhibition or activation. The classical molecules explored in this thesis with known biological activities were the flavonoid, *trans*-chalcone a plant NP and garlic-derived allicin. The chalcone flavonoids are known to have antimicrobial properties against a number of microbes including MRSA. However, not many microbial enzyme targets that interact with the chalcones have been investigated.

In this thesis, sortase A, a promising target in Gram positive bacteria was observed to be inhibited by *trans*-chalcone *in vitro*. A study of the interaction between *S. mutans* SrtA and *trans*-chalcone led to the identification of a Michael adduct formed between the active site cysteine residue of the enzyme and the *β*-carbon of *trans*-chalcone using MS analysis. This modification contributes to the observed inhibition of the enzyme by *trans*-chalcone *in vitro*. From this study, we were also able to report the first crystal structure of the H139A mutant of *S. mutans* SrtA, which gave insights into the binding pocket that the diaryl molecule sits in the active site using molecular modelling techniques. Such data has implications for future design and identification of other chalcone flavonoids with inhibitory activities against the sortase family of enzymes.

Another interesting covalent adduct was also identified in this body of work between allicin and the peroxidase BCP from *Burkholderia* as described in chapter 5. The thiosulphinate derived from garlic extracts reacted with a catalytic cysteine residue of BCP to produce an allythio derivative. This result was very surprising, as we were able to show that crude garlic extract could be analysed with a biological enzyme probe (BCP) to fish out the active molecule from such a complex mixture. From this project we were also able to report for the first time that allicin and allicin-containing garlic extracts possess bactericidal activities against the Bcc. These findings merit further investigation as adjuncts to existing antibiotics to fight infections caused by Bcc in CF patients.

Protein MS was utilisied extensively in this thesis which highlights the importance of this emerging tool in NP research. Modern MS techniques can facilitate the identification of covalently-bound adducts formed between various NPs and protein targets and also provide data that can support the elucidation and verification of mechanisms.[207] This was very evident in determining the molecular basis for the inhibition of *S. mutans* SrtA by *trans*-chalcone. Plant NPs also presents a challenge in characterizing the bioactive molecules which are usually present in small quantities and in complex mixtures. However, the use of protein MS is an ideal analytical tool to identify bioactive molecules, using protein targets as biological probes. This approach was utilized in identifying allicin as the main bioactive component of garlic extract that interacted with BCP from *Burkholderia*. Additionally, the technique can be used to identify protein targets from complex proteomes using NPs as chemical probes.

In summary, the work presented in this thesis has shown that NP investigations remain an active and diverse field of academic research as analytical techniques continue to evolve and whole-genome sequencing become more accessible. There are many lessons to be learnt from studies of classical plant NPs as well as novel compounds that are yet to be explored. With the need to combat infectious diseases, the future will likely see a renewed commitment to NPs and a drive towards maximizing genomics-driven NP

discovery in an effort to identify bioactive molecules, to unravel the mechanism of new enzymes and gain insights into NP biosynthesis.[208]

References

1. Walsh, C.T. & Fischbach, M.A. Natural products version 2.0: connecting genes to molecules. *J Am Chem Soc* **132**, 2469-93 (2010).
2. Davies, J. & Ryan, K.S. Introducing the Parvome: Bioactive Compounds in the Microbial World. *ACS Chem. Bio.* **7**, 252-259 (2011).
3. Li, J.W. & Vederas, J.C. Drug discovery and natural products: end of an era or an endless frontier? *Science* **325**, 161-5 (2009).
4. Schulze, C.J. et al. "Function-first" lead discovery: mode of action profiling of natural product libraries using image-based screening. *Chem Biol* **20**, 285-95 (2013).
5. Morrison, K.C. & Hergenrother, P.J. Natural products as starting points for the synthesis of complex and diverse compounds. *Nat Prod Rep* **31**, 6-14 (2013).
6. Challis, G.L. Genome mining for novel natural product discovery. *J Med Chem* **51**, 2618-28 (2008).
7. Cragg, G.M. & Newman, D.J. Natural products: a continuing source of novel drug leads. *Biochim Biophys Acta* **1830**, 3670-95 (2013).
8. Rudd, B.A. & Hopwood, D.A. Genetics of actinorhodin biosynthesis by *Streptomyces coelicolor* A3(2). *J Gen Microbiol* **114**, 35-43 (1979).
9. Caffrey, P. et al. An acyl-carrier-protein-thioesterase domain from the 6-deoxyerythronolide B synthase of *Saccharopolyspora erythraea*. High-level production, purification and characterisation in *Escherichia coli*. *Eur J Biochem* **195**, 823-30 (1991).
10. Cortes, J., Haydock, S.F., Roberts, G.A., Bevitt, D.J. & Leadlay, P.F. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature* **348**, 176-8 (1990).
11. Meier, J.L. & Burkart, M.D. The chemical biology of modular biosynthetic enzymes. *Chem Soc Rev* **38**, 2012-2045 (2009).
12. Bachmann, B.O., Van Lanen, S.G. & Baltz, R.H. Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? *J Ind Microbiol Biotechnol* **41**, 175-84 (2014).
13. Butler, M.S. & Buss, A.D. Natural products--the future scaffolds for novel antibiotics? *Biochem Pharmacol* **71**, 919-29 (2006).
14. Bush, K. et al. Tackling antibiotic resistance. *Nat Rev Microbiol* **9**, 894-6 (2011).
15. Cassir, N., Rolain, J.M. & Brouqui, P. A new strategy to fight antimicrobial resistance: the revival of old antibiotics. *Front Microbiol* **5**, 551 (2014).
16. Coenye, T., Vandamme, P., Govan, J. & Lipauma, J. Taxanomy and Indentification of the *Burkholderia cepacia* complex. *J Clin Microbiol* **39**, 3427-3436 (2001).
17. Govan, J., Brown, A. & Jones, A. Evolving epidemiology of pseudomonas aeroginosa and the *Burkholderia cepacia* complex in cystic fibrosis lung infection. *Future Microbiol* **2**, 153-164 (2007).
18. Greenberg, D. et al. Recurrent *Burkholderia* infection in patients with chronic granulomatous disease. *Clin Infect Dis* **48**, 1577-1579 (2009).

19. Nzula, S., Vandamme, P. & Govan, J. Influence of taxonomic status on the in vitro antimicrobial suseptibility of the *Burkholderia cepacia* complex *J Antimicrob Chemother* **50**, 265-269 (2002).
20. Grimwood, K., Kidd, T.J. & Tweed, M. Successful treatment of cepacia syndrome. *J Cyst Fibros* **8**, 291-3 (2009).
21. Waters, V. & Ratjen, F. Multidrug-resistant organisms in cystic fibrosis: management and. *Expert Rev Anti Infect Ther.* **4**, 807-819 (2006).
22. LiPuma, J. The changing microbial epidemiology in cystic fibrosis. *Clin Microbiol Rev* **23**, 299-323 (2010).
23. Tullis, D.E. et al. Inhaled aztreonam for chronic *Burkholderia* infection in cystic fibrosis: a placebo-controlled trial. *J Cyst Fibros* **13**, 296-305 (2014).
24. Balfour-Lynn, I.M. At last, *Burkholderia* spp. is one of the inclusion criteria--a negative (but published) randomised controlled trial. *J Cyst Fibros* **13**, 241-2 (2014).
25. Schwan, W. et al. Screening a mushroom extract library for activity against *Acinetobacter baumannii*. *Ann Clin Microbiol Antimicrob.* **9** (2010).
26. Mahenthiralingam, E. et al. Enacyloxins are products of an unusual hybrid modular polyketide synthase encoded by a cryptic *Burkholderia ambifaria* Genomic Island. *Chem Biol* **18**, 665-77 (2011).
27. Mil-Homens, D., Bernardes, N. & Fialho, A. The antibacterial properties of docosahexaenoic omega-3 fatty acid against the. *FEMS Microbiol Lett.* **328**, 61-69 (2012).
28. Papaleo, M.C. et al. Sponge-associated microbial Antarctic communities exhibiting antimicrobial activity against *Burkholderia cepacia* complex bacteria. *Biotechnol Adv* **30**, 272-93 (2012).
29. Fischbach, M. & Walsh, C. Antibiotics for Emerging Pathogens. *Science* **325**, 1089-1093 (2009).
30. Li, J. & Vederas, J. Drug Discovery and natural products. *Science* **325**, 161-165 (2009).
31. Williamson, N.R. et al. Biosynthesis of the red antibiotic, prodigiosin, in *Serratia*: identification of a novel 2-methyl-3-n-amyl-pyrrole (MAP) assembly pathway, definition of the terminal condensing enzyme, and implications for undecylprodigiosin biosynthesis in *Streptomyces*. *Mol Microbiol* **56**, 971-89 (2005).
32. Cerdeno, A.M., Bibb, M.J. & Challis, G.L. Analysis of the prodiginine biosynthesis gene cluster of *Streptomyces coelicolor* A3(2): new mechanisms for chain initiation and termination in modular multienzymes. *Chem Biol* **8**, 817-29 (2001).
33. Mo, S. et al. Elucidation of the *Streptomyces coelicolor* pathway to 2-undecylpyrrole, a key intermediate in undecylprodiginine and streptorubin B biosynthesis. *Chem Biol* **15**, 137-48 (2008).
34. Franks, A. et al. Isolation and structure elucidation of a novel yellow pigment from the marine bacterium *Pseudoalteromonas tunicata*. *Molecules* **10**, 1286-91 (2005).
35. Carbone, M. et al. A new cytotoxic tambjamine alkaloid from the Azorean nudibranch *Tambja ceutae*. *Bioorg Med Chem Lett* **20**, 2668-70 (2010).

36. Carte, B. & Faulkner, D.J. Defensive metabolites from three nembrothid nudibranchs. *J Org Chem* **48**, 2314-2318 (1983).
37. Kojiri, K., Nakajima, S., Suzuki, H., Okura, A. & Suda, H. A new antitumor substance, BE-18591, produced by a *Streptomycete*. I. Fermentation, isolation, physico-chemical and biological properties. *J Antibiot* **46**, 1799-803 (1993).
38. Nakajima, S., Kojiri, K. & Suda, H. A new antitumor substance, BE-18591, produced by a *Streptomycete*. II. Structure determination. *J Antibiot* **46**, 1894-6 (1993).
39. Burke, C., Thomas, T., Egan, S. & Kjelleberg, S. The use of functional genomics for the identification of a gene cluster encoding for the biosynthesis of an antifungal tambjamine in the marine bacterium *Pseudoalteromonas tunicata*. *Environ Microbiol* **9**, 814-8 (2007).
40. Franks, A. et al. Inhibition of fungal colonization by *Pseudoalteromonas tunicata* provides a competitive advantage during surface colonization. *Appl Environ Microbiol* **72**, 6079-87 (2006).
41. Iglesias Hernandez, P. et al. Tambjamine alkaloids and related synthetic analogs: efficient transmembrane anion transporters. *Chem Commun* **48**, 1556-8 (2012).
42. Hernando, E., Soto-Cerrato, V., Cortes-Arroyo, S., Perez-Tomas, R. & Quesada, R. Transmembrane anion transport and cytotoxicity of synthetic tambjamine analogs. *Org Biomol Chem* **12**, 1771-8 (2014).
43. Bowman, J.P. Bioactive compound synthetic capacity and ecological significance of marine bacterial genus *Pseudoalteromonas*. *Mar Drugs* **5**, 220-41 (2007).
44. Burke, C., Thomas, T., Egan, S. & Kjelleberg, S. The use of functional genomics for the identification of a gene cluster. *Environmental Microbiology* **9**, 814 (2007).
45. Whicher, J.R. et al. Structure and function of the RedJ protein, a thioesterase from the prodiginine biosynthetic pathway in *Streptomyces coelicolor*. *J Biol Chem* **286**, 22558-69 (2011).
46. Garneau-Tsodikova, S., Dorrestein, P.C., Kelleher, N.L. & Walsh, C.T. Protein assembly line components in prodigiosin biosynthesis: characterization of PigA,G,H,I,J. *J Am Chem Soc* **128**, 12600-1 (2006).
47. Walsh, C.T., Garneau-Tsodikova, S. & Howard-Jones, A.R. Biological formation of pyrroles: nature's logic and enzymatic machinery. *Nat Prod Rep* **23**, 517-31 (2006).
48. Stanley, A.E., Walton, L.J., Kourdi Zerikly, M., Corre, C. & Challis, G.L. Elucidation of the *Streptomyces coelicolor* pathway to 4-methoxy-2,2'-bipyrrole-5-carboxaldehyde, an intermediate in prodiginine biosynthesis. *Chem Commun (Camb)*, 3981-3 (2006).
49. Merrill, A.H., Jr. De novo sphingolipid biosynthesis: a necessary, but dangerous, pathway. *J Biol Chem* **277**, 25843-6 (2002).
50. Alexeev, D. et al. The crystal structure of 8-amino-7-oxononanoate synthase: a bacterial PLP-dependent, acyl-CoA-condensing enzyme. *J Mol Biol* **284**, 401-19 (1998).
51. Kellmann, R. et al. Biosynthetic intermediate analysis and functional homology reveal a saxitoxin gene cluster in cyanobacteria. *Appl Environ Microbiol* **74**, 4044-53 (2008).

52. Gerber, R., Lou, L. & Du, L. A PLP-dependent polyketide chain releasing mechanism in the biosynthesis of mycotoxin fumonisins in *Fusarium verticillioides*. *J Am Chem Soc* **131**, 3148-9 (2009).
53. Jahan, N. et al. Insights into the biosynthesis of the *Vibrio cholerae* major autoinducer CAI-1 from the crystal structure of the PLP-dependent enzyme CqsA. *J Mol Biol* **392**, 763-73 (2009).
54. Spirig, T. et al. The Legionella autoinducer synthase LqsA produces an alpha-hydroxyketone signaling molecule. *J Biol Chem* **283**, 18113-23 (2008).
55. Schmidt, A. et al. Three-dimensional structure of 2-amino-3-ketobutyrate CoA ligase from *Escherichia coli* complexed with a PLP-substrate intermediate: inferred reaction mechanism. *Biochemistry* **40**, 5151-60 (2001).
56. Yard, B.A. et al. The Structure of Serine Palmitoyltransferase; Gateway to Sphingolipid Biosynthesis. *J Mol Biol* **370**, 870-886 (2007).
57. Raman, M.C., Johnson, K.A., Clarke, D.J., Naismith, J.H. & Campopiano, D.J. The serine palmitoyltransferase from *Sphingomonas wittichii* RW1: An interesting link to an unusual acyl carrier protein. *Biopolymers* **93**, 811-22 (2010).
58. Raman, M.C. et al. The external aldimine form of serine palmitoyltransferase: structure, kinetic and spectroscopic analysis of the wild-type enzyme and HSAN1 mutant mimics. *J Biol Chem* **284**, 17328-39 (2009).
59. Astner, I. et al. Crystal structure of 5-aminolevulinate synthase, the first enzyme of heme biosynthesis, and its link to XLSA in humans. *Embo J* **24**, 3166-77 (2005).
60. Kelly, R.C. et al. The *Vibrio cholerae* quorum-sensing autoinducer CAI-1: analysis of the biosynthetic enzyme CqsA. *Nat Chem Biol* **5**, 891-5 (2009).
61. Kerbarh, O., Campopiano, D.J. & Baxter, R.L. Mechanism of [small alpha]-oxoamine synthases: identification of the intermediate Claisen product in the 8-amino-7-oxononanoate synthase reaction. *Chem Commun*, **7**, 60-62 (2006).
62. Zhang, W., Bolla, M.L., Kahne, D. & Walsh, C.T. A three enzyme pathway for 2-amino-3-hydroxycyclopent-2-enone formation and incorporation in natural product biosynthesis. *J Am Chem Soc* **132**, 6402-11 (2010).
63. Wei, Y., Perez, L.J., Ng, W.L., Semmelhack, M.F. & Bassler, B.L. Mechanism of *Vibrio cholerae* autoinducer-1 biosynthesis. *ACS Chem Biol* **6**, 356-65 (2011).
64. Hornung, C. et al. The *Janthinobacterium sp*. HH01 genome encodes a homologue of the *V. cholerae* CqsA and *L. pneumophila* LqsA autoinducer synthases. *PLoS One* **8**, e55045 (2013).
65. Eliot, A.C. & Kirsch, J.F. Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu Rev Biochem* **73**, 383-415 (2004).
66. Dunathan, H.C. Conformation and reaction specificity in pyridoxal phosphate enzymes. *Proc Natl Acad Sci U S A* **55**, 712-6 (1966).
67. Arthur, C. et al. The Malonyl Transferase Activity of Type II PKS ACPs. *Chemistry and Biology* **13**, 587-596 (2006).
68. Lai, J.R., Koglin, A. & Walsh, C.T. Carrier Protein Structure and Recognition in Polyketide and Nonribosomal Peptide Biosynthesis. *Biochemistry* **50**, 14869-14879 (2006).
69. Crosby, J. & Crump, M. The structural role of the carrier protein. *Nat Prod Rep* **29**, 1111-37 (2012).

70. Chan, D. & Vogel, H. Current understanding of fatty acid biosynthesis and the acyl carreir protein. *Biochem J.* **43**, 1-19 (2010).
71. Halavaty, A. et al. Structural Charcterisation and comparison of three ACPs from pathogenic bacteria. *Biol Crystallogr* **68**, 1359-1370 (2012).
72. Connolly, K. et al. Sortase from *Staphylococcus aureus* does not contain a thiolate-imodazolium ion pair. *J Biol Chem* **278**, 34061-34065 (2003).
73. Lai, J.R., Koglin, A. & Walsh, C.T. Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis. *Biochemistry* **45**, 14869-79 (2006).
74. Lim, J. et al. Rigidifying Acyl Carrier Protein Domain in Iterative Type I PKS CalE8 Does Not Affect Its Function. *Biophys J* **103**, 1037-1044 (2012).
75. Mercer, A.C. & Burkart, M.D. The ubiquitous carrier protein--a window to metabolite biosynthesis. *Nat Prod Rep* **24**, 750-73 (2007).
76. Cronan, J.E. The chain-flipping mechanism of ACP (acyl carrier protein)-dependent enzymes appears universal. *Biochem J* **460**, 157-63 (2014).
77. Parris, K.D. et al. Crystal structures of substrate binding to *Bacillus subtilis* holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites. *Structure* **8**, 883-95 (2000).
78. Rafi, S. et al. Structure of acyl carrier protein bound to FabI, the FASII enoyl reductase from *Escherichia coli. J Biol Chem* **281**, 39285-93 (2006).
79. Nguyen, C. et al. Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. *Nature* **505**, 427-31 (2014).
80. Masoudi, A., Raetz, C.R., Zhou, P. & Pemble, C.W.t. Chasing acyl carrier protein through a catalytic cycle of lipid A production. *Nature* **505**, 422-6 (2014).
81. Cryle, M.J. & Schlichting, I. Structural insights from a P450 Carrier Protein complex reveal how specificity is achieved in the P450(BioI) ACP complex. *Proc Natl Acad Sci U S A* **105**, 15696-701 (2008).
82. Agarwal, V., Lin, S., Lukk, T., Nair, S.K. & Cronan, J.E. Structure of the enzyme-acyl carrier protein (ACP) substrate gatekeeper complex required for biotin synthesis. *Proc Natl Acad Sci U S A* **109**, 17406-11 (2012).
83. Dholwani, K.K., Saluja, A.K., Gupta, A.R. & Shah, D.R. A review on plant-derived natural products and their analogs with anti-tumor activity. *Indian J Pharmacol* **40**, 49-58 (2008).
84. Veitch, N.C. & Grayer, R.J. Flavonoids and their glycosides, including anthocyanins. *Nat Prod Rep* **28**, 1626-95 (2011).
85. Kumar, S. & Pandey, A.K. Chemistry and biological activities of flavonoids: an overview. *ScientificWorldJournal* **2013**, 162750 (2013).
86. Batovska, D.I. & Todorova, I.T. Trends in utilization of the pharmacological potential of chalcones. *Curr Clin Pharmacol* **5**, 1-29 (2010).
87. Orlikova, B., Tasdemir, D., Golais, F., Dicato, M. & Diederich, M. Dietary chalcones with chemopreventive and chemotherapeutic potential. *Genes Nutr* **6**, 125-47 (2011).
88. Rahman, M. Chalcone: A Valuable Insight into the Recent Advances and Potential Pharmacological Activites. *Chemical Sciences Journal* **11**, 1-16 (2011).

89. Block, E. Garlic and Other Alliums: The Lore and the Science (The Royal Society of Chemistry, Cambridge, UK, 2010).
90. Small, L.D., Bailey, J.H. & Cavallito, C.J. Comparison of Some Properties of Thiolsulfonates and Thiolsulfinates. *J Am Chem Soc* **71**, 3565-3566 (1949).
91. Small, L.D., Bailey, J.H. & Cavallito, C.J. Alkyl thiolsulfinates. *J Am Chem Soc* **69**, 1710-3 (1947).
92. Cavallito, C.J., Bailey, J.H. & Buck, J.S. The Antibacterial Principle of *Allium sativum*. III. Its Precursor and "Essential Oil of Garlic"1. *J Am Chem Soc* **67**, 1032-1033 (1945).
93. Cavallito, C.J., Buck, J.S. & Suter, C.M. Allicin, the antibacterial principle of *Allium sativum* II Determination of the chemical structure. *J Am Chem Soc* **66**, 1952-1954 (1944).
94. Cavallito, C.J. & Bailey, J.H. Allicin, the antibacterial principle of *Allium sativum* I Isolation, physical properties and antibacterial action. *J Am Chem Soc* **66**, 1950-1951 (1944).
95. Kyung, K.H. Antimicrobial properties of allium species. *Curr Opin Biotechnol* **23**, 142-7 (2012).
96. Lawson, L.D., Wood, S.G. & Hughes, B.G. HPLC analysis of allicin and other thiosulfinates in garlic clove homogenates. *Planta Med* **57**, 263-70 (1991).
97. Kaschula, C.H., Hunter, R. & Parker, M.I. Garlic-derived anticancer agents: structure and biological activity of ajoene. *Biofactors* **36**, 78-85 (2010).
98. Jakobsen, T.H. et al. Ajoene, a sulfur-rich molecule from garlic, inhibits genes controlled by quorum sensing. *Antimicrob Agents Chemother* **56**, 2314-25 (2012).
99. Borlinghaus, J., et. al. Molecules **19**, 12591-12618 (2014).
100. Ankri, S. & Mirelman, D. Antimicrobial properties of allicin from garlic. *Microbes Infect* **1**, 125-9 (1999).
101. Fujisawa, H., Watanabe, K., Suma, K., Origuchi, K. & Matsufuji, H. Antimicrobial Potential of Garlic-Derived Allicin and Its Cancellation by Sulfydryl Compounds. *Biosc Biotechnol Biochem* **73**, 1948-1955 (2009).
102. Davis, L.E., Shen, J. & Royer, R.E. In vitro synergism of concentrated *Allium sativum* extract and amphotericin B against *Cryptococcus neoformans*. *Planta Med* **60**, 546-9 (1994).
103. Shen, J., Davis, L.E., Wallace, J.M., Cai, Y. & Lawson, L.D. Enhanced diallyl trisulfide has in vitro synergy with amphotericin B against *Cryptococcus neoformans*. *Planta Med* **62**, 415-8 (1996).
104. Didry, N., Dubreuil, L. & Pinkas, M. Antimicrobial activity of naphtoquinones and Allium extracts combined with antibiotics. *Pharm Acta Helv* **67**, 148-51 (1992).
105. Sohn, D.W. et al. Anti-inflammatory and antimicrobial effects of garlic and synergistic effect between garlic and ciprofloxacin in a chronic bacterial prostatitis rat model. *Int J Antimicrob Agents* **34**, 215-9 (2009).
106. Miron, T., Rabinkov, A., Mirelman, D., Wilchek, M. & Weiner, L. The mode of action of allicin: its ready permeability through phospholipid membranes may contribute to its biological activity. *Biochim Biophys Acta* **1463**, 20-30 (2000).

107. Cutler, R.R. & Wilson, P. Antibacterial activity of a new, stable, aqueous extract of allicin against methicillin-resistant *Staphylococcus aureus*. *Br J Biomed Sci* **61**, 71-4 (2004).
108. Feldberg, R., Chang, S., Kotik, A., Nadler, M. & Neuwirth, Z. In vitro mechanism of inhibition of bacterial cell growth by allicin. *Antimicrob Agents Chemother* **32**, 1763-1768 (1988).
109. Focke, M., Feld, A. & Lichtenthaler, H. Allicin, a naturally occuring antibiotic from garlic, specifically inhibits acetyl-CoA synthase. *FEBS Letters* **261**, 106-108 (1990).
110. Rabinkov, A. et al. The mode of action of allicin: trapping of radicals and interaction with thiol containing proteins. *Biochim Biophys Acta* **1379**, 233-44 (1998).
111. Waag, T. et al. Allicin and derivatives are cysteine protease inhibitors with parasitic activity. *Bioorg Med Chem Lett* **20**, 5541-5543 (2010).
112. Ankri, S., Miron, T., Rabinkov, A., Wilchek, M. & Mirelman, D. Allicin from garlic strongly inhibits cysteine proteinases and cytopathic effects of *Entamoeba histolytica*. *Antimicrob Agents Chemother* **41**, 2286-8 (1997).
113. Marraffini, L.A., Dedent, A.C. & Schneewind, O. Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol Mol Biol Rev* **70**, 192-221 (2006).
114. Spirig, T., Weiner, E.M. & Clubb, R.T. Sortase enzymes in Gram-positive bacteria. *Mol Microbiol* **82**, 1044-59 (2011).
115. Clancy, K.W., Melvin, J.A. & McCafferty, D.G. Sortase transpeptidases: insights into mechanism, substrate specificity, and inhibition. *Biopolymers* **94**, 385-96 (2010).
116. Maresso, A.W. & Schneewind, O. Sortase as a target of anti-infective therapy. *Pharmacol Rev* **60**, 128-41 (2008).
117. Mazmanian, S.K., Liu, G., Ton-That, H. & Schneewind, O. *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **285**, 760-3 (1999).
118. Hendrickx, A.P.A., Budzik, J.M., Oh, S.-Y. & Schneewind, O. Architects at the bacterial surface - sortases and the assembly of pili with isopeptide bonds. *Nat Rev Microbiol* **9**, 166-176 (2011).
119. Mazmanian, S.K., Ton-That, H., Su, K. & Schneewind, O. An iron-regulated sortase anchors a class of surface protein during *Staphylococcus aureus* pathogenesis. *Proc Natl Acad Sci U S A* **99**, 2293-8 (2002).
120. Zong, Y., Mazmanian, S.K., Schneewind, O. & Narayana, S.V. The structure of sortase B, a cysteine transpeptidase that tethers surface protein to the *Staphylococcus aureus* cell wall. *Structure* **12**, 105-12 (2004).
121. Marraffini, L.A. & Schneewind, O. Targeting proteins to the cell wall of sporulating *Bacillus anthracis*. *Mol Microbiol* **62**, 1402-17 (2006).
122. Nobbs, A.H., Lamont, R.J. & Jenkinson, H.F. *Streptococcus* adherence and colonization. *Microbiol Mol Biol Rev* **73**, 407-50, (2009).
123. Igarashi, T. Deletion in sortase gene of *Streptococcus mutans* Ingbritt. *Oral Microbiol Immunol* **19**, 210-3 (2004).

124. Igarashi, T., Asaga, E. & Goto, N. Roles of *Streptococcus mutans* dextranase anchored to the cell wall by sortase. *Oral Microbiol Immunol* **19**, 102-5 (2004).
125. Jenkinson, H.F. & Demuth, D.R. Structure, function and immunogenicity of streptococcal antigen I/II polypeptides. *Mol Microbiol* **23**, 183-90 (1997).
126. Hamada, S. & Slade, H.D. Biology, immunology, and cariogenicity of *Streptococcus mutans*. *Microbiol Rev* **44**, 331-84 (1980).
127. Sato, Y., Yamamoto, Y. & Kizaki, H. Cloning and sequence analysis of the gbpC gene encoding a novel glucan-binding protein of *Streptococcus mutans*. *Infect Immun* **65**, 668-75 (1997).
128. Igarashi, T., Asaga, E. & Goto, N. The sortase of *Streptococcus mutans* mediates cell wall anchoring of a surface protein antigen. *Oral Microbiol Immunol* **18**, 266-9 (2003).
129. Igarashi, T., Asaga, E., Sato, Y. & Goto, N. Inactivation of srtA gene of *Streptococcus mutans* inhibits dextran-dependent aggregation by glucan-binding protein C. *Oral Microbiol Immunol* **19**, 57-60 (2004).
130. Crowley, P.J., Brady, L.J., Michalek, S.M. & Bleiweis, A.S. Virulence of a spaP mutant of *Streptococcus mutans* in a gnotobiotic rat model. *Infect Immun* **67**, 1201-6 (1999).
131. Loesche, W.J. Role of *Streptococcus mutans* in human dental decay. *Microbiol Rev* **50**, 353-80 (1986).
132. Colby, S.M., Whiting, G.C., Tao, L. & Russell, R.R. Insertional inactivation of the *Streptococcus mutans* dexA (dextranase) gene results in altered adherence and dextran catabolism. *Microbiology* **141** (Pt 11), 2929-36 (1995).
133. Bolken, T.C. et al. Inactivation of the srtA gene in *Streptococcus gordonii* inhibits cell wall anchoring of surface proteins and decreases in vitro and in vivo adhesion. *Infect Immun* **69**, 75-80 (2001).
134. Zong, Y.N., Bice, T.W., Ton-That, H., Schneewind, O. & Narayana, S.V.L. Crystal structures of *Staphylococcus aureus* sortase A and its substrate complex. *J Biol Chem* **279**, 31383-31389 (2004).
135. Ilangovan, U., Ton-That, H., Iwahara, J., Schneewind, O. & Clubb, R.T. Structure of sortase, the transpeptidase that anchors proteins to the cell wall of *Staphylococcus aureus*. *Proc Natl Acad Sci U S A* **98**, 6056-61 (2001).
136. Race, P.R. et al. Crystal structure of *Streptococcus pyogenes* sortase A: implications for sortase mechanism. *J Biol Chem* **284**, 6924-33 (2009).
137. Ton-That, H., Liu, G., Mazmanian, S.K., Faull, K.F. & Schneewind, O. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proc Natl Acad Sci U S A* **96**, 12424-12429 (1999).
138. Ton-That, H. & Schneewind, O. Anchor structure of staphylococcal surface proteins. IV. Inhibitors of the cell wall sorting reaction. *J Biol Chem* **274**, 24316-20 (1999).
139. Ton-That, H., Mazmanian, S.K., Faull, K.F. & Schneewind, O. Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Sortase catalyzed in vitro transpeptidation reaction using LPXTG peptide and NH(2)-Gly(3) substrates. *J Biol Chem* **275**, 9876-81 (2000).

140. Jacobitz, A.W. et al. Structural and computational studies of the *Staphylococcus aureus* sortase B-substrate complex reveal a substrate-stabilized oxyanion hole. *J Biol Chem* **289**, 8891-902 (2014).
141. Mao, H., Hart, S.A., Schink, A. & Pollok, B.A. Sortase-mediated protein ligation: a new method for protein engineering. *J Am Chem Soc* **126**, 2670-1 (2004).
142. Williamson, D.J., Fascione, M.A., Webb, M.E. & Turnbull, W.B. Efficient N-terminal labeling of proteins by use of sortase. *Angew Chem Int Ed Engl* **51**, 9377-80 (2012).
143. Antos, J.M. et al. Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. *J Am Chem Soc* **131**, 10800-1 (2009).
144. Bellucci, J.J., Bhattacharyya, J. & Chilkoti, A. A Noncanonical Function of Sortase Enables Site-Specific Conjugation of Small Molecules to Lysine Residues in Proteins. *Angew Chem Int Ed Engl* **54**, 441-5 (2014).
145. Hu, P., Huang, P. & Chen, M.W. Curcumin reduces *Streptococcus mutans* biofilm formation by inhibiting sortase A activity. *Arch Oral Biol* **58**, 1343-8 (2013).
146. Chenna, B.C. et al. Identification of novel inhibitors of bacterial surface enzyme *Staphylococcus aureus* Sortase A. *Bioorg Med Chem Lett* **18**, 380-5 (2008).
147. Maresso, A.W. et al. Activation of inhibitors by sortase triggers irreversible modification of the active site. *J Biol Chem* **282**, 23129-39 (2007).
148. Frankel, B.A., Bentley, M., Kruger, R.G. & McCafferty, D.G. Vinyl sulfones: inhibitors of SrtA, a transpeptidase required for cell wall protein anchoring and virulence in *Staphylococcus aureus. J Am Chem Soc* **126**, 3404-5 (2004).
149. Scott, C.J. et al. Irreversible inhibition of the bacterial cysteine protease-transpeptidase sortase (SrtA) by substrate-derived affinity labels. *Biochem J* **366**, 953-8 (2002).
150. Park, B.S. et al. *Curcuma longa L*. constituents inhibit sortase A and *Staphylococcus aureus* cell adhesion to fibronectin. *J Agric Food Chem* **53**, 9005-9009 (2005).
151. Kim, S.H. et al. Inhibition of sortase, a bacterial surface protein anchoring transpeptidase, by beta-sitosterol-3-O-glucopyranoside from *Fritillaria verticillata*. *Biosci Biotechnol Biochem* **67**, 2477-9 (2003).
152. Oh, I. et al. *In Vitro* Sortase A Inhibitory and Antimicrobial Activity of Flavonoids Isolated from the Roots of Sophora flavescens. *Archives of Pharmacal Research* **34**, 217-222 (2011).
153. Kang, S.S., Kim, J.G., Lee, T.H. & Oh, K.B. Flavonols inhibit sortases and sortase-mediated *Staphylococcus aureus* clumping to fibrinogen. *Biol Pharm Bull* **29**, 1751-5 (2006).
154. Huang, P., Hu, P., Zhou, S.Y., Li, Q. & Chen, W.M. Morin Inhibits Sortase A and Subsequent Biofilm Formation in *Streptococcus mutans. Curr Microbiol* **68**, 47-52 (2013).
155. Hu, P., Huang, P. & Chen, W.M. Curcumin inhibits the Sortase A activity of the *Streptococcus mutans* UA159. *Appl Biochem Biotechnol* **171**, 396-402 (2013).
156. Liu, H. & Naismith, J.H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol* **8**, 91 (2008).

157. Liu, H. & Naismith, J.H. A simple and efficient expression and purification system using two newly constructed vectors. *Protein Expr Purif* **63**, 102-11 (2009).
158. Clarke, D.J. et al. Subdivision of the bacterioferritin comigratory protein family of bacterial peroxiredoxins based on catalytic activity. *Biochemistry* **49**, 1319-30 (2010).
159. Fujisawa, H., Suma, K., Origuchi, K., Kumagai, H. & Seki, T. Biological and chemical stability of garlic-derived allicin. *J Agric Food Chem* **56**, 4229-4235 (2008).
160. Lawson, L. & Wang, Z.J. Low Allicin Release from Garlic Supplements: a Major problem Due to the Sensitivities of Allinase Activity. *J Agric Food Chem* **49**, 2592-2599 (2001).
161. Jakowicz, J. Principles of Fluorescence Spectroscopy (Springer, Baltimore, 2006).
162. Rupp, B. Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology. *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*, 1-809 (2010).
163. Zhang, R. et al. Structures of sortase B from *Staphylococcus aureus* and *Bacillus anthracis* reveal catalytic amino acid triad in the active site. *Structure* **12**, 1147-56 (2004).
164. Stein, N. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J Appl Crystallogr* **41**, 641-643 (2008).
165. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
166. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).
167. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
168. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).
169. Schrodinger, LLC. (2010).
170. Exterkate, R.A., Crielaard, W. & Ten Cate, J.M. Different response to amine fluoride by *Streptococcus mutans* and polymicrobial biofilms in a novel high-throughput active attachment model. *Caries Res* **44**, 372-9 (2010).
171. NCCLS. 1-182 (National Committee for Clinical Laboratory Standards, Wayne, PA, USA, 1997).
172. Xu, G.Y. et al. Solution structure of *B. subtilis* acyl carrier protein. *Structure* **9**, 277-87 (2001).
173. Martinez, M.A. et al. A novel role of malonyl-ACP in lipid homeostasis. *Biochemistry* **49**, 3161-7 (2010).
174. Mozzarelli, A. & Bettati, S. Exploring the pyridoxal 5'-phosphate-dependent enzymes. *Chem Rec* **6**, 275-87 (2006).
175. Yang, Z.R., Thomson, R., McNeil, P. & Esnouf, R.M. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* **21**, 3369-76 (2005).

176. G, W. xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* **43**, 186-190 (2010).
177. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-3 (1994).
178. Ikushiro, H. et al. Structural insights into the enzymatic mechanism of serine palmitoyltransferase from *Sphingobacterium multivorum*. *J Biochem* **146**, 549-62 (2009).
179. Schneewind, O. & Missiakas, D.M. Protein secretion and surface display in Gram-positive bacteria. *Philos Trans R Soc Lond B Biol Sci* **367**, 1123-39 (2012).
180. Lee, S.F. & Boran, T.L. Roles of sortase in surface expression of the major protein adhesin P1, saliva-induced aggregation and adherence, and cariogenicity of *Streptococcus mutans*. *Infect Immun* **71**, 676-81 (2003).
181. Ton-That, H., Mazmanian, S.K., Alksne, L. & Schneewind, O. Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Cysteine 184 and histidine 120 of sortase form a thiolate-imidazolium ion pair for catalysis. *J Biol Chem* **277**, 7447-52 (2002).
182. Matayoshi, E.D., Wang, G.T., Krafft, G.A. & Erickson, J. Novel fluorogenic substrates for assaying retroviral proteases by resonance energy transfer. *Science* **247**, 954-8 (1990).
183. Zong, Y., Bice, T., Ton-That, H., Schneewind, O. & Narayana, S. Crystal structures of *Staphylococcus aureus* sortase A and its substrate complex. *J Biol Chem.* **279**, 31383-9 (2004).
184. Lemos, J.A., Quivey, R.G., Jr., Koo, H. & Abranches, J. *Microbiology* **159** 436-45 (2013).
185. Naganawa, R. et al. Inhibition of Microbial Growth by Ajoene, a Sulfur-Containing Compound Derived from Garlic. *Applied and Envirnmental Microbiology* **62**, 4238-4242 (1996).
186. Smyth, A. et al. Garlic as an Inhibitor of *Pseudomonas aeruginosa Pediatric Pulmonology* **45**, 356-362 (2010).
187. Jakobsen, T. et al. Ajoene, a sulfur-rich molecule from garlic, inhibits genes controlled by quorum. *Antimicrob Agents Chemother* **56**, 2314-2325 (2012).
188. Holmes, A., Govan, J. & Goldstein, R. Agricultural use of *Burkholderia (Pseudomonas) cepacia*: a threat to human health. *Emerging Infectious Diseases* **4** (1998).
189. Sullivan, L.A. & Mahenthiralingam, E. Biotechnological potential within the genus *Burkholderia*. *Letters in Applied Microbiology* **41**, 8-11 (2005).
190. Burkholder, W. Sour skin, a bacterial rot of onion bulbs. *Phytopathology* **40**, 115-117 (1950).
191. Horsley, A., Webb, K., Bright-Thomas, R., Govan, J. & Jones, A. Can early *Burkholderia cepacia* complex infection in cystic fibrosis be eradicated with antibiotic therapy? *Front Cell Infect Microbiol* **1**, 18 (2011).
192. Horsley, A. & Jones, A.M. Antibiotic treatment for *Burkholderia cepacia* complex in people with cystic fibrosis experiencing a pulmonary exacerbation. *Cochrane Database Syst Rev* **10**, Cd009529 (2012).

193. Block, E. Garlic and Other Alliums: The Lore and the Science (The Royal Society of Chemistry, Cambridge, UK, 2010).
194. Miron, T., Listowsky, I. & Wilchek, M. Reaction mechanisms of allicin and allyl-mixed disulfides with proteins and small thiol molecules. *Eur J Med Chem* **45**, 1912-8 (2010).
195. Lawson, L.D. & Gardner, C.D. Composition, stability, and bioavailability of garlic products used in a clinical trial. *J Agric Food Chem* **53**, 6254-61 (2005).
196. Lawson, L.D., Wang, Z.J. & Hughes, B.G. Identification and HPLC quantitation of the sulfides and dialk(en)yl thiosulfinates in commercial garlic products. *Planta Med* **57**, 363-70 (1991).
197. Mahenthiralingam, E. et al. Diagnostically and experimentally useful panel of strains from the *Burkholderia*. *J Clin Microbiol* **38**, 910-913 (2000).
198. Geoghegan, K.F. et al. Spontaneous alpha-N-6-phosphogluconoylation of a "His tag" in *Escherichia coli*: the cause of extra mass of 258 or 178 Da in fusion proteins. *Anal Biochem* **267**, 169-84 (1999).
199. Taylor, K. et al. Covalent dimer species of beta-defensin Defr1 display potent antimicrobial. *Antimicrob Agents Chemother* **51**, 1719-1724 (2007).
200. Loutet, S.A. & Valvano, M.A. Extreme antimicrobial peptide and polymyxin B resistance in the genus *Burkholderia*. *Front Cell Infect Microbiol* **1**, 6 (2011).
201. Schwan, W.R. et al. Screening a mushroom extract library for activity against Acinetobacter baumannii and *Burkholderia cepacia* and the identification of a compound with anti-Burkholderia activity. *Ann Clin Microbiol Antimicrob* **9**, 4 (2010).
202. Mil-Homens, D., Bernardes, N. & Fialho, A.M. The antibacterial properties of docosahexaenoic omega-3 fatty acid against the cystic fibrosis multiresistant pathogen *Burkholderia cenocepacia*. *FEMS Microbiol Lett* **328**, 61-9 (2012).
203. Ashry, N.A., Gameil, N.M. & Suddek, G.M. Modulation of cyclophosphamide-induced early lung injury by allicin. *Pharm Biol* **51**, 806-11 (2013).
204. Banerjee, S.K. & Maulik, S.K. Effect of garlic on cardiovascular disorders: a review. *Nutr J* **1**, 4 (2002).
205. Zhang, L., Wang, E., Chen, F., Yan, H. & Yuan, Y. Potential protective effects of oral administration of allicin on acrylamide-induced toxicity in male mice. *Food Funct* **4**, 1229-36 (2013).
206. Luo, Y., Cobb, R.E. & Zhao, H. Recent advances in natural product discovery. *Curr Opin Biotechnol* **30c**, 230-237 (2014).
207. Bouslimani, A., Sanchez, L.M., Garg, N. & Dorrestein, P.C. Mass spectrometry of natural products: current, emerging and future technologies. *Nat Prod Rep* **31**, 718-29 (2014).
208. Milshteyn, A., Schneider, J.S. & Brady, S.F. Mining the metabiome: identifying novel natural products from microbial communities. *Chem Biol* **21**, 1211-23 (2014).
209. Larkin, M.A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-8 (2007).

210. Gouet, P., Robert, X. & Courcelle, E. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* **31**, 3320-3 (2003).

Appendices

Appendix 1

Figure A1: AOS family of enzymes. Fourteen AOSs have been identified to date. They carry our condensation between a specific amino acid and a thioester substrate.

Enzyme	Abbreviation	Pathway	Amino Acid Substrate	Thioester Substrate	Condensation Product
2-amino-3-ketobutyrate CoA ligase	KBL	Threonine degradation	Glycine	Acetyl-CoA	α-ammio-β-ketobutyrate (AKB)
5-aminolevulinate synthase	ALAS	Heme	Glycine	Succinyl-CoA	5-aminolevulinic acid (ALA)
8-amino-7-oxonanoate synthase	AONS	Biotin	L-alanine	Pimelyol-CoA	(7S)-8-amino-7-oxononanoate (AON)
Serine palmitoyltransferase	SPT	Sphingolipid	L-serine	Palmitoyl-CoA/ACP	3-ketodihydrosphingosine (KDS)
Cholerea CAI-1 synthase	CqsA	Quorum-sensing autoinducer	*S*-amino butyrate/ SAM	Decanoyl-CoA	3-hydroxytridecan-4-one/ cholera autoinducer-1 (CAI-1)
Legionella LA-1 synthase	LqsA	Quorum-sensing autoinducer	SAM	?	3-hydroxypentadecan-4-one/Legionella autoinducer-1 (LAI-1)
Alanine transferase	JqsA	Quorum-sensing autoinducer	?	?	JAI-1
Glycyl transerfase	RedN	Prodiginine	Glycine	β-ketomyristoyl-ACP	4-keto-2-undecylpyrroline
Seryl transferase	RedL	Prodiginine	L-serine	pyrrolyl-β-ketoacyl-S-ACP	4-hydroxy-2-2'-bipyrrole-5-methanol (HBM)
Seryl transferase	PigH	Prodigiosin	L-serine	pyrrolyl-β-ketoacyl-S-ACP	4-hydroxy-2-2'-bipyrrole-5-methanol (HBM)
Fumonisin ORF8	Fum8p	Fumonisin	L-alanine	Steroyl-ACP	Sphingosine intermediate
Arginine transferase	SxtA	Saxitosin	Arginine	Propionyl-ACP	4-amino-3-oxo-guanidinoheptane
Seryl transferase	TamD	Tambjamines	L-serine	pyrrolyl-β-ketoacyl-S-ACP	4-hydroxy-2-2'-bipyrrole-5-methanol (HBM)
Orf34	Orf34	2-amino-3-hydroxycyclopent-2-enone polyketide(C_5N)	Glycine	Succinyl-CoA	ALA

Appendix 2

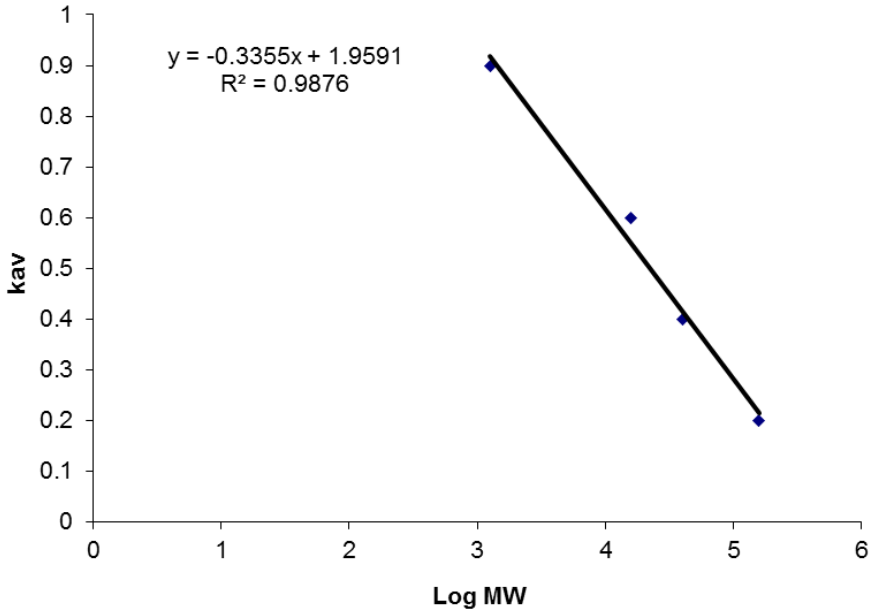


Figure A2: Calibration curve for Superdex S200 16/60 column (GE Healthcare). Gel filtration standard (Biorad) contains thyroglobulin (670 kDa), γ-globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa) and vitamin B_{12} (1.35 kDa). Buffer composition: 25 mM Tris-HCl, 500 mM NaCl, 10% glycerol, pH 8. Flow Rate: 1 mL/min.

Appendix 3

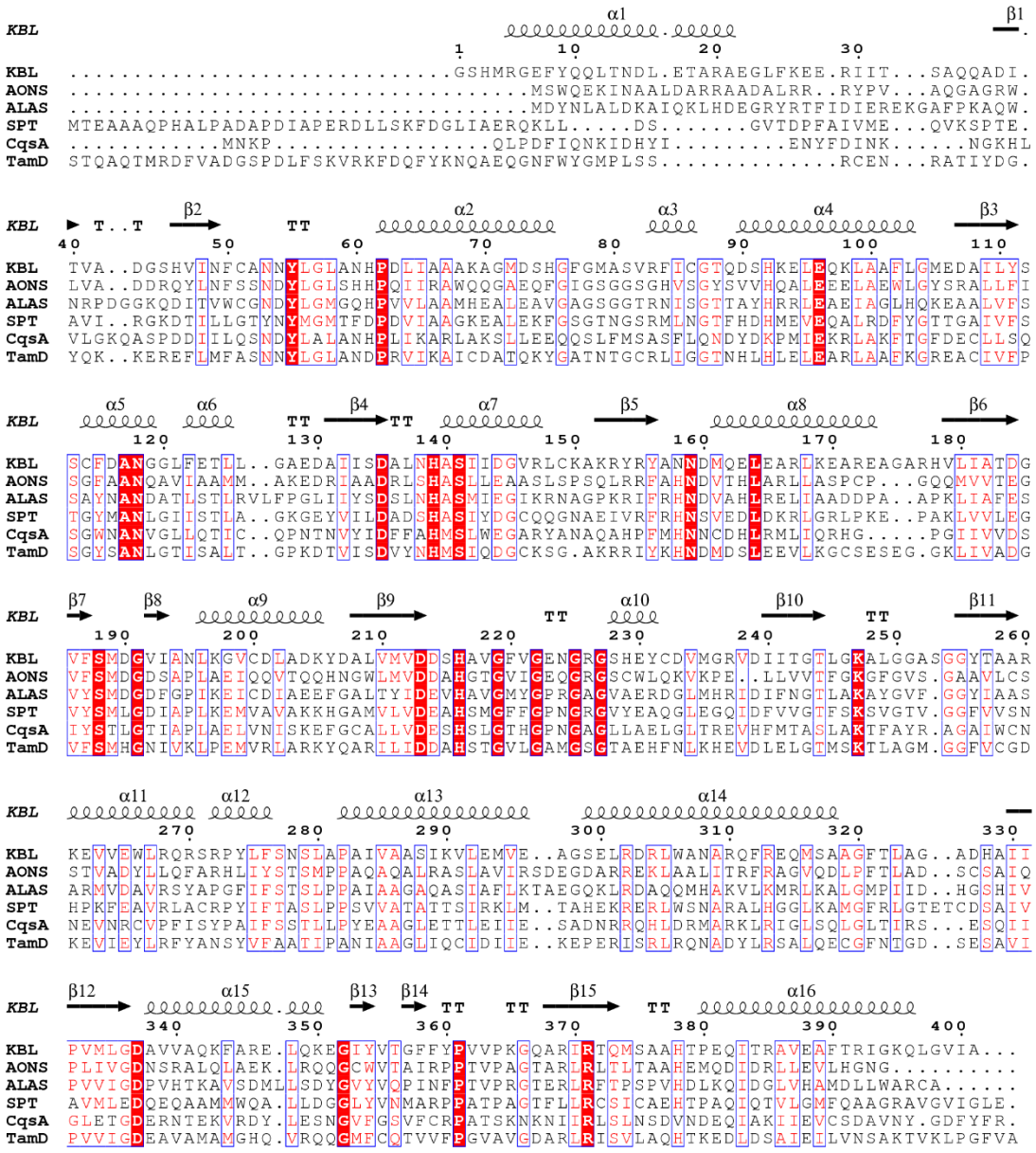


Figure A3: Sequence alignment of TamD predicted seryltransferase domain with members of the AOS family. TamD (A4C5W2), SPT (2JG2; Q93UV0), KBL (1FC4; B5YWB8), AONS (1DJE; C4ZXU7), ALAS (2BWN; P18079)) and CqsA (2WK8; C3NXB8). The secondary structure (coil for α-helix and arrow for β-strand) of KBL is shown above the sequences. Residues involved in PLP binding are indicated with blue triangle. This figure was produced using ClustalW2[209] and ESPript.[210] *The TamD sequence in this alignment starts at residue 117. PDB codes are in black and UniProt codes in blue.

Appendix 4

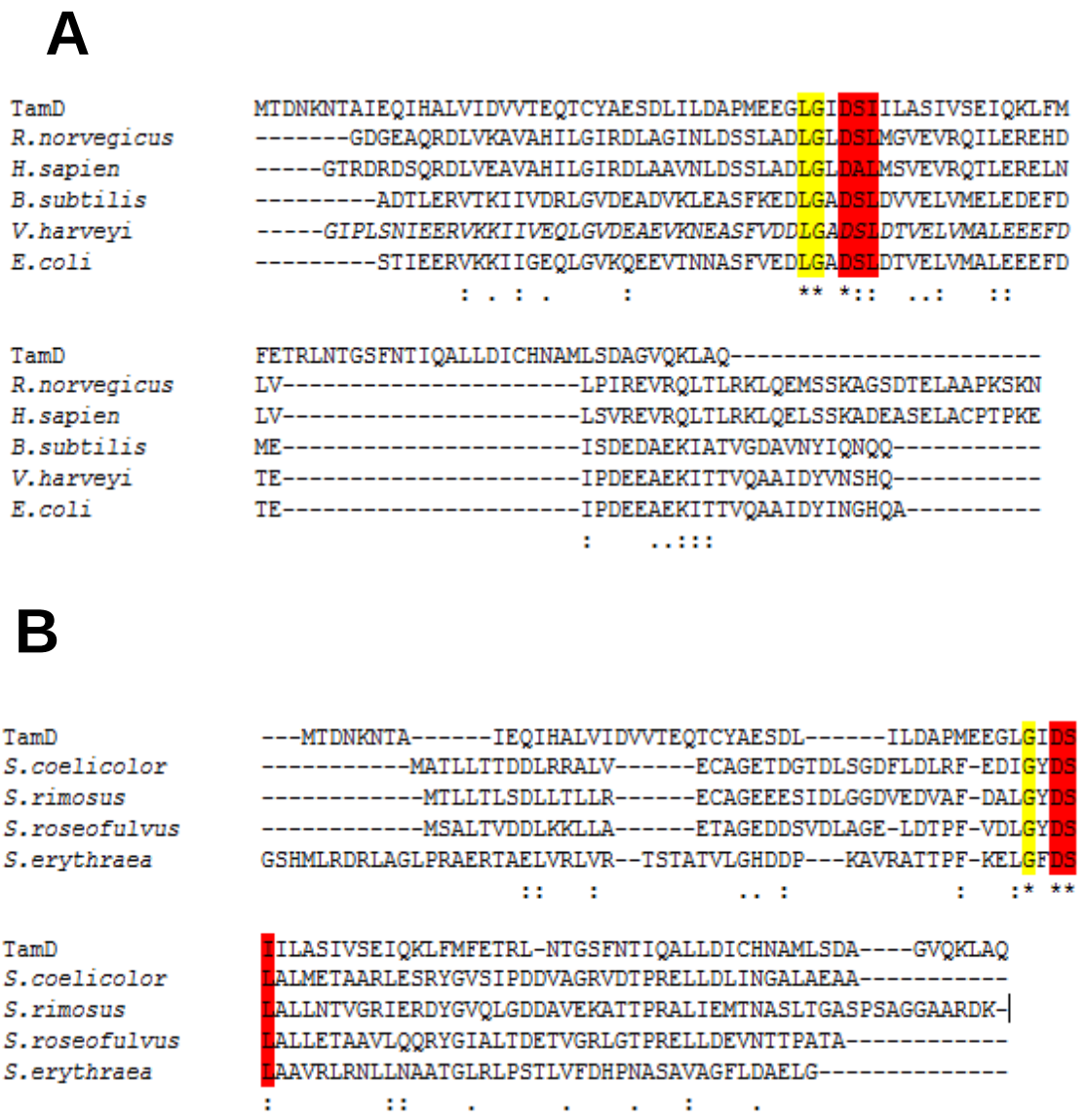


Figure A4: Sequence alignment of TamD-ACP domain with ACPs. (A) Alignment with FAS ACPs: *R. norvegicus* (1N8L; P12785), *H. sapien* (2CG5; P49327), *B. subtilis* (1HY8; E0TTR0), *V. harveyi* (2L0Q; DOXCG4), *E. coli* (1T8K; A0A0B5JX78). (B) Alignment with PKS ACPs: *S. coelicolor* (2K0Y; Q02054), *S. rimosus* (1NQ4; Q3S8R2), *S. roseofulvus* (1ORS; Q54996), *S. erythraea* (2JU1; Q03131). The conserved DSL/I motif is highlighted in red and residues with high sequence identity are highlighted in yellow which undergoes PTM by the addition of a phosphopantetheinyl prosthetic group. PDB codes are in black and UniProt codes in blue.

Appendix 5

1 MTDNKNTAIE QIHALVIDVV TEQTCYAESD LILDAPMEEG LGIDSIILAS IVSEIQKLFM FETRLNTGSF NTIQALLDIC
81 HNAMLSDAGV QKLAQLGLAA APQAVCVSSQ PEPEQRSTQA QTMRDFVADG SPDLFSKVRK FDQFYKNQAE QGNFWYGMPL
161 SSRCENRATI YDGYQKKERE FLMFASNNYL GLANDPRVIK AICDATQKYG ATNTGCRLIG GTNHLHLELE ARLAAFKGRE
241 ACIVFPSGYS ANLGTISALT GPKDTVISDV YNHMSIQDGC KLSGAKRRIY KHNDMDSLEE VLKGCSESEG GKLIVADGVF
321 SMHGNIVKLP EMVRLARKYQ ARILIDDAHS TGVLGAMGSG TAEHFNLKHE VDLELGTMSK TLAGMGGFVC GDKEVIEYLR
401 FYANSYVFAA TIPANIAAGL IQCIDIIEKE PERISRLRQN ADYLRSALQE CGFNTGDSES AVIPVVIGDE AVAMAMGHQV
481 RQQGMFCQTV VFPGVAVGDA RLRISVLAQH TKEDLDSAIE ILVNSAKTVK LPGFVALEHH HHHH

Figure A5: **Peptide map from TamD digest by enoproteinase LysC**. The data was SNAPed and searched using MS Fit and error of 20 ppm. The sequences in red were observed peptides (79%). The matched peptides covered 79% of the protein sequence i.e. 432 of 544 amino acids.

Appendix 6

Nmol/asym	Matthews Coeff	% solvent	P(4.98)	P(tot)
1	12.56	90.21	0.00	0.00
2	6.28	80.42	0.00	0.00
3	4.19	70.63	0.03	0.02
4	3.14	60.84	0.18	0.14
5	2.51	51.05	0.48	0.45
6	2.09	41.26	0.29	0.37
7	1.79	31.47	0.01	0.02
8	1.57	21.69	0.00	0.00
9	1.40	11.90	0.00	0.00
10	1.26	2.11	0.00	0.00

Figure A6: Matthew's calculations for TamD. The number of molecules expected in the ASU is between 4 and 6 based on Matthew's number calculations. This is consistent with Matthew's coefficient between 2.09 and 3.14 and solvent content of 41.3 and 60.8 %. The actual number of molecules in the ASU is 4, which is consistent with a higher solvent content and consequently a weaker diffracting crystal.

Appendix 7

Ncyc	Rfact	Rfree
0	0.4533	0.4775
1	0.4533	0.4793
2	0.4520	0.4785
3	0.4507	0.4779
4	0.4485	0.4761
5	0.4448	0.4725
6	0.4378	0.4666
7	0.4190	0.4607
8	0.4161	0.4589
9	0.4038	0.4584
10	0.3838	0.4482

Figure A7: Refinement of TamD model. Ten refinement cycles were performed in REFMAC illustrating improvement of the R-factor and R-free values.

Appendix 8

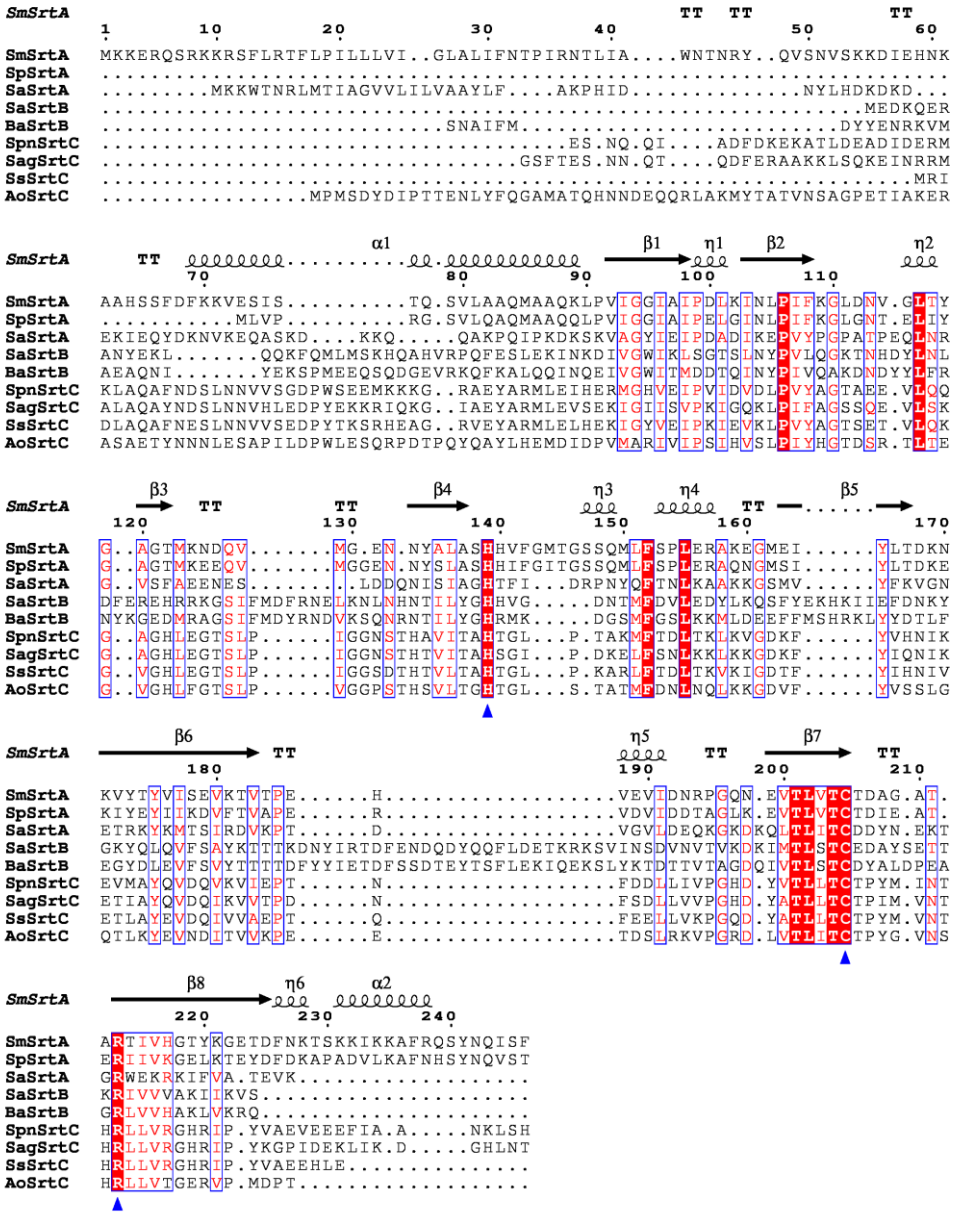


Figure A8: Sequence alignment of sortases. Identical residues are highlighted in red, similar residues are highlighted with a blue box. Catalytic triad is indicated with a blue Δ. *S. mutans* (SmSrtA) (4TQX; I6L907), *S. pyogenes* (SpSrtA) (3FN5; Q1JGU0) *S. aureus* (SaSrtA) (1T2P; D9RJF4) *S. aureus* (SaSrtB) (1NG5; A0A090LW34), .B. *anthracis* (BaSrtB) (2OQW; A0A0B5YM13) *S. pnemoniae* (SpnSrtC) (2W1J; G6LSC3), *S. agalactiae* (SagSrtC) (3O0P; SAS8MMM0), *S. suis* (SsSrtC) (3RE9; A4VY75), *A. oris* (AoSrtC) (2XWG; Q0Z952).

Appendix 9

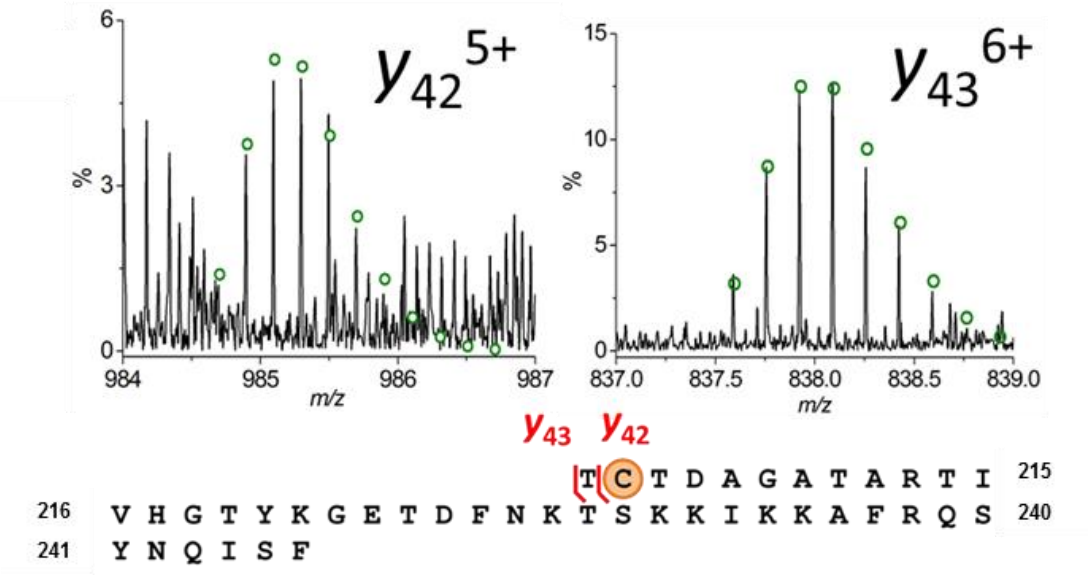


Figure A9: Mass spectrum of CID fragments ions used to assign site of modification in SmSrtA. The isotopic distributions of the y_{42}^{5+} ($[C_{207}H_{333}N_{60}O_{64}S_1(C_{15}H_{12}O)]^{5+}$) and y_{43}^{6+} ($[C_{211}H_{341}N_{61}O_{66}S_1(C_{15}H_{12}O)]^{6+}$) fragment ions allowed for the assignment the site of modification as Cys205. The simulated isotope patterns are shown as scatterplots in green.

Appendix 10

Gene sequence pET28a/TamD WT (*Pseudoalteromonas tunicata*)

*cat**atg***acagacaataaaaacaccgctattgagcaaatacacgccctagttatcgacgtagtaacagagca
aacgtgttatgccgagtcggatttaattttagatgcgccgatggaagaaggcttggggatagactccatta
tccttgcttctatcgtcagtgaaattcaaaaattgtttatgtttgagacccgtctcaatactggcagtttt
aataccattcaagcattactcgacatttgtcacaatgcgatgctatcagacgcaggagtgcaaaaactggc
acaattaggacttgcagcagcaccacaagctgtttgtgtaagttcgcagccagagcctgaacagcgttcaa
ctcaggcacaaacaatgcgagattttgtcgcagatggtagccctgacttatttagtaaagtgcgtaagttt
gaccagttttataaaaatcaggctgagcaaggtaacttttggtacggcatgccacttagctccagatgtga
aaatcgagcgactatttatgatggctatcagaaaaaagaacgtgaattcttaatgtttgcctcgaataatt
atttggggttagctaacgacccgcgggtcatcaaagcaatctgtgatgctacgcaaaaatacggtgcaaca
aatacaggttgtcgtctgattggtggcactaatcatttgcaccttgaactggaagcacgtttggcagcgtt
taaaggtcgcgaagcctgtattgttttcccctctggttattcggctaaccttggtacgatttctgcgttaa
ctggtccaaaagacactgtgatttcagatgtttataatcacatgagtattcaagatggttgtaagttatca
ggtgcaaaacgccgtatttacaaacataacgatatggattcgttagaagaagtactaaagggttgcagtga
gtccgaaggtggtaagctaattgttgctgatggtgtgttcagcatgcacggtaacattgtaaaactgccag
aaatggtgcgtctggcccgtaaatatcaagcacgtattttgattgatgatgcccattctactggggtgtta
ggggcaatggggtcaggtactgctgagcatttcaacctcaaacatgaagtcgatcttgaattaggcaccat
gagtaaaacgctggctggtatgggggggatttgtctgtggtgacaaagaagtaattgagtatttacgttttt
acgcaaattcttatgtgtttgctgcaactattcctgcaaatatcgcggcgggtttaattcaatgtattgat
attattgaaaaagagccagagagaattagtcgtttacgccaaaatgcagattacctgcgcagtgcattaca
agaatgcggttttaataccggtgacagtgaaagtgccgtcattcctgtggttattggcgatgaagctgtag
caatggcaatgggccatcaagttcgtcagcaaggcatgttctgtcaaacggtagtatttccgggtgttgca
gttggtgatgcacgtttacggatcagcgttctggctcagcatacaaaagaagacctcgatagtgccattga
gattttagtgaactcagctaaaacagtgaaattacccggttttgtggca*ctcgag*caccaccaccaccacc
actga

TamD Protein Sequence

```
  1 MTDNKNTAIE QIHALVIDVV TEQTCYAESD LILDAPMEEG LGIDSIILAS IVSEIQKLFM  60
 61 FETRLNTGSF NTIQALLDIC HNAMLSDAGV QKLAQLGLAA APQAVCVSSQ PEPEQRSTQA 120
121 QTMRDFVADG SPDLFSKVRK FDQFYKNQAE QGNFWYGMPL SSRCENRATI YDGYQKKERE 180
181 FLMFASNNYL GLANDPRVIK AICDATQKYG ATNTGCRLIG GTNHLHLELE ARLAAFKGRE 240
241 ACIVFPSGYS ANLGTISALT GPKDTVISDV YNHMSIQDGC KLSGAKRRIY KHNDMDSLEE 300
301 VLKGCSESEG GKLIVADGVF SMHGNIVKLP EMVRLARKYQ ARILIDDAHS TGVLGAMGSG 360
361 TAEHFNLKHE VDLELGTMSK TLAGMGGFVC GDKEVIEYLR FYANSYVFAA TIPANIAAGL 420
421 IQCIDIIEKE PERISRLRQN ADYLRSALQE CGFNTGDSES AVIPVVIGDE AVAMAMGHQV 480
481 RQQGMFCQTV VFPGVAVGDA RLRISVLAQH TKEDLDSAIE ILVNSAKTVK LPGFVALEHH 540
541 HHHH
```

Appendix 11

Gene sequence pET22b/SrtA-N40 WT (*Streptococcus mutans*)

ccatgggcagcagcgcttggaataccaatagatatcaggtttctaatgttagcaagaaagatattgaacac
aacaaggctgcccattcttcctttgattttaaaaaggtggaatctatcagtactcaatcggtactggcagc
gcaaatggctgctcagaagcttcctgtaattggcggaattgccattccagacttaaaaatcaatttaccaa
tcttcaaaggattagataatgttggcttaacatatggtgctggaacgatgaaaaatgaccaagtcatggga
gaaaataattatgctcttgctagccatcatgtttttggtatgaccggatcttcacagatgctcttttcacc
tttagaacgtgcaaaagaaggcatggaaatttatctgactgataaaaataaggtttatacttatgttatta
gtgaagtgaaaactgtcacacctgaacatgtagaagttattgacaatcggccgggacaaaatgaagttact
ttggtcacttgtacagatgcgggggcgactgccagaacaattgttcatggcacatataaggggggaaactga
ttttaataagacttccaaaaagataaaaaaaagcttttaggcagtcctataatcaaatatcattttaactcg
ag

SrtA-N40 Protein Sequence

41	AWNTNRYQVS	NVSKKDIEHN	KAAHSSFDFK	KVESISTQSV	LAAQMAAQKL	PVIGGIAIPD	100
101	LKINLPIFKG	LDNVGLTYGA	GTMKNDQVMG	ENNYALASHH	VFGMTGSSQM	LFSPLERAKE	160
161	GMEIYLTDKN	KVYTYVISEV	KTVTPEHVEV	IDNRPGQNEV	TLVTCTDAGA	TARTIVHGTY	220
221	KGETDFNKTS	KKIKKAFRQS	YNQISF				

Publications